Universidad Nacional Autónoma de México

Facultad de Ciencias

Modelos de sociofísica y econofísica en evaluación de desorden e índices de desigualdad en estructuras conservativas jerárquicas

T E S I S

QUE PARA OBTENER EL TÍTULO DE:
FÍSICO

PRESENTA:
ERIC HERNÁNDEZ RAMÍREZ

DIRECTOR DE TESIS:
DR. MARCELO DEL CASTILLO MUSSOT

2014

MÉXICO., D.F.

Hoja de datos del jurado

1. Datos del alumno
Hernández
Ramírez
Eric
56 10 30 99
Universidad Nacional Autónoma de México
Facultad de Ciencias
Física
096335123

2. Datos del tutor
Dr.
Marcelo
Del Castillo
Mussot

3. Datos del sinodal 1
Dr.
Pedro Eduardo
Miramontes
Vidal

4. Datos del sinodal 2
Dr.
Wolf Luis
Mochán
Backal

5. Datos del sinodal 3
Dr.
Jorge Humberto
Arce
Rincón

6. Datos del sinodal 4
Dr.
Alfredo Elmer
De la Lama
García

7. Datos del trabajo escrito
Modelos de sociofísica y econofísica en evaluación de desorden e índices de desigualdad en estructuras conservativas jerárquicas
73 pag
2014

Modelos de sociofísica y econofísica en evaluación de desorden e índices de desigualdad en estructuras conservativas jerárquicas .

Eric Hernádez Ramírez.

Índice general

Capítulo 1

Introducción

Resulta que los físicos, son perfectamente adecuados para invadir las demás disciplinas,... Los físicos tienden a verse como los señores de la jungla académica, considerando sus propios métodos por encima de los otros y celosos guardianes de su propio territorio. Sus alter egos son cercanos a los de un animal oportunista, felices de pedir prestadas ideas de cualquier lugar si les parecen que les pueden ser útiles, y encantados de entrometerse en problemas ajenos. Por irritante que sea esta actitud para todos, la llegada de los físicos a un área de investigación distinta a la suya a menudo es un presagio de un período emocionante y de grandes descubrimientos. Los matemáticos hacen lo mismo de vez en cuando, pero nadie desciende con tanta furia y en tan gran número como los físicos, como manada hambrienta, y bajo el influjo de la adrenalina por el olor de un nuevo problema.

Duncan J. Watts [Six Degrees The Science of a connected age]

Todos sabemos que ni la riqueza ni los ingresos se distribuyen de manera equitativa en una población; unos cuantos muy ricos y muchos pobres. Sin embargo, la afirmación de que la riqueza y el ingreso se distribuye de manera desigual plantea muchas preguntas: ¿como es la distribución de la riqueza en las sociedades?, ¿es una característica particular de algunos países o esta presente en todos?, ¿qué mecanismo la produce?, entre otras. Uno de los primeros en investigar este tipo de preguntas fue Vilfredo Pareto (sociólogo, economista y filósofo italiano 1896-97),quien usando modelos matemáticos y evidencia estadística busco regularidades en los datos y logro determinar la forma de la curva de distribución del ingreso; la cual en su parte final estaba bien descrita por una ley de potencia, sin embargo, en esa época los datos son insuficientes para caracterizarla completamente, como lo señala Pareto en su obra. Desde un punto de vista actual el principal aporte de Pareto fue mostrar que esta distribución varia muy poco independientemente de la población y la época. Y además descubrió que esta curva, no corresponde a una distribución Gaussiana, donde la conservación y adquisición de la riqueza dependen solo del azar.

A simple vista uno podría pensar que el estudio de la desigualdad en una sociedad no es un terreno para la Física, sin embargo, en los últimos años los físicos han avanzado

desde los temas tradicionales a nuevas disciplinas llevando consigo su forma de pensar y sus métodos. En este sentido, fue el físico teórico Eugene Stanley quien en 1995 introduce el termino Econofísica en la conferencia Dynamics of Complex Systems en Kolkata. En su correspondiente articulo, Stanley [poner ref], presenta un manifiesto del nuevo campo bajo el argumento "el comportamiento de un gran número de seres humanos (medida por ejemplo, por los índices económicos) podría ajustarse a los análogos de las leyes de escala que han resultado útiles para describir sistemas compuestos por un gran número de objetos inanimados". Debido a la naturaleza estadística de la Econofísica debemos señalar que esta no aplica literalmente las leyes de la física, tal como las leyes de Newton o la mecánica cuántica, a los humanos sino que, para analizar las propiedades de estos sistemas complejos, compuestos de un gran numero de seres humanos, utiliza métodos desarrollados para la física estadística, distinguiéndose de la estadística y probabilidad matemática por su enfoque, métodos y resultados. En general podemos decir que existen dos formas de abordar estos problemas, el primero es analizar los datos estadísticos de los que se dispone y el segundo diseñar modelos teóricos que permiten reproducir el comportamiento de estos datos. Este nuevo campo interdisciplinario ha crecido en varias direcciones: teoría macroeconómica (distribución de la riqueza), micro estructura de los mercados financieros (modelación del libro de pedidos), econometría de las crisis y burbujas financieras, entre otras [Ecnophysics:Empirical facts]. Es importante mencionar, que todos estos avances han sido posibles gracias al desarrollo de la computación y que actualmente están disponibles grandes bases de datos con información estadística y económica de varios países en el mundo, es sobre esta información que se pueden obtener resultados empíricos importantes que permiten aventurarse a desarrollar modelos teóricos que puedan dar luz sobre las propiedades de estos sistemas.

Otro campo muy cercano a la Econofísica es Sociofísica término que a sido reivindicado por el físico teórico Serge Galam [art Sociophysics] desde principios de la década 1980. Actualmente la Sociofísica es reconocida como un campo vinculado a la Física Estadística que estudia e intenta explicar un rango mas amplio de temas sociales, que cada vez son mas numeroso, entre los cuales destacan : las redes sociales, la evolución del lenguaje, la dinámica de poblaciones, la propagación de las epidemias, el terrorismo, las votaciones, la formación de coaliciones y las dinámicas de opinión; esta última ha producido una gran cantidad de trabajos de investigación. En los últimos años la globalización económica ha hecho que los fenómenos colectivos religiosos, políticos, económicos y sociales sean el foco de atención de los acontecimientos mundiales y que la idea de señalar una conección entre los fenómenos colectivos físicos y el comportamiento colectivo de sistemas humanos parezca algo natural. Esto plantea varias preguntas fundamentales una de ellas, ¿se comportan los humanos como átomos?, a lo que Galam responde: ?si en ciertas partes del comportamiento colectivo social y político? y ?no en otros aspectos de la conducta humana?. Como todo nuevo campo la Sociofísica necesita nuevos conceptos y la tentación de tomarlos de la Física Estadística es constante y peligrosa puesto que asignar una teoría física construida para una realidad física a una realidad social no pasaría de una bonita metáfora que seria incapaz de predecir y ,peor aun, podría convertirse en una interpretación peligrosa de una Teoría Social. En este

sentido Galam nos dice ?nuestra tarea consiste en pedir prestados de la Física técnicas y conceptos que puedan ser utilizados para construir una teoría de la conducta colectiva, pero dentro de las limitaciones especificas de la realidad psicosocial. El peligro para el físico es permanecer en la Física, usando una terminología social y un formalismo físico. La contribución de la Física por lo tanto debería limitarse a las directrices para el modelado de la realidad social sin que esto implique que nuestro programa sea menos ambicioso sino por el contrario, abre perspectivas mas solidas para hacer frente de manera eficiente al actual proceso de globalización?. Por lo tanto es importante recalcar que la Sociofísica no pretende llegar a una descripción exacta de un grupo humano, sino mas bien tratar de explicar fenómenos humanos que son lo bastante complicados como para asumir cualquier cosa como verdadera.

Después, de señalar cual es el papel del físico al abordar temas sociales, podemos plantear el objetivo de este trabajo, el cual consiste en desarrollar un índice de desigualdad que nos permita medir que tan jerárquica es una estructura. Uno de los principales referentes de un índice de desigualdad es Corrado Gini a quien le debemos el desarrollo del coeficiente que lleva su nombre en 1912 [Poner cita], dicho índice es una medida de la desigualdad en los ingresos de una población, actualmente es uno los índices mas utilizados tanto por economistas como gobiernos y entidades internacionales para medir la desigualdad, debido a su facilidad de calculo y de interpretación,y en el cual nos apoyaremos para desarrollar nuestra idea. Sin embargo existe otra característica de la desigualdad entre los individuos que forman parte de una sociedad, que no se reduce a que cantidad de ingreso o riqueza que se posee y que se ha comenzado a reflejar en los modelos de distribución de ingreso; estas distribuciones muestran que la sociedad esta formada por dos conjuntos bien diferenciadas por el tipo de distribución que obedecen; la mayoría de la población (97-99 %) pertenecientes a la clase baja esta caracterizada por la distribución exponencial de Boltzmann- Gibbs, mientras que la clase alta (1-3 % de la población) tiene una distribución de ley de potencia de Pareto [Two-class Structure]. Este hecho nos sirve de motivación para plantear la desigualdad posicional apoyándonos en el concepto de red . En cierta forma, todos sabemos que nuestras relaciones familiares, de amistad, el puesto laboral o si soy miembro de una organización o no, me permite influir en mi entorno y ser influenciado por él, este hecho establece que ocupamos un lugar en una red independientemente de que la identifique o no, donde mis relaciones locales son solo una parte de una red global que involucra a muchos mas individuos. En palabras de Nicholas [conectados], las dinámicas que se establecen dentro de estas redes sociales pueden reforzar espectacularmente dos tipos diferentes de desigualdad: la desigualdad situacional (personas que se encuentran mejor desde un punto de vista socioeconómico) y la desigualdad posicional (personas que disfrutan de mejor posición dependiendo del lugar en que se encuentren en el interior de la red). Estas diferencias entre los individuos por el lugar que ocupan nos llevan a dos concepto importantes: el de red como ya hemos mencionado y el de jerarquía. El estudio de las redes reales en las últimas décadas ha tenido un gran auge como un campo interdisciplinario, donde los físicos han hecho importantes aportes, por el momento no nos detendremos en este punto puesto que le dedicaremos un capítulo en el presente trabajo. Por otro lado,el concepto de jerarquía y estratificación ha sido un tema debatido

por los sociólogos por muchos años. Sin entrar en detalles los distintos enfoques clásicos (marxista, weberiana y funcionalista) de la estratificación social comparten el hecho de introducir conceptos para el análisis de diferenciación y jerarquización como son: clase social, estatus, posición de mercado, estructura ocupacional, entre otras, estos enfoques aunque tienen distintas perspectivas del análisis comparten la idea de la centralidad del trabajo en los procesos de formación y diferenciación social [Vea CEPAl].Por lo tanto podríamos decir que existen dos componentes principales de la desigualdad uno socioeconomico y otro estructural y este último, se refleja principalmente en la estructura laboral. En esta dirección existe un importante trabajo que investiga la relación entre los índices socioeconomicos y los de prestigio basados en la clasificación de las ocupaciones por su prestigio en la sociedad.[Hauser Socioeconomic index].

Capítulo 2

Indice de desigualdad de Gini

En este capítulo abordamos una de las medidas mas utilizadas para determinar el grado de desigualdad en la distribución de ingreso. Una de las principales características de esta medida que hace de ella un punto de referencia para los debates y toma de decisiones en torno a la desigualdad en las sociedades es su fácil determinación, así como, una representación gráfica que permite una interpretación sencilla. Esta medida nos servirá de base para la construcción del índice de desigualdad en redes jerárquicas que proponemos.

2.1. Desigualdad

El ser humano al ser una especie social constantemente esta interaccionando de distintas formas, muchas de estas interacciones son complejas y se dan entre los individuos que constituyen una sociedad. A medida que una sociedad alcanza cierto grado de complejidad muestra desigualdad entre sus miembros. Uno de los factores que llegan a contribuir de manera importante a esto, es la estratificación jerárquica de las clases sociales [4], la cual contribuye a la distribución desigual en los ingresos, no siendo esta última la única desigualdad que se puede manifestar. Probablemente uno de los primeros en investigar la desigualdad del ingreso de manera cuantitativa fue Vilfredo Pareto (1848-1923) quien analizó datos de distintos países entre ellos se encontraban Inglaterra, Perú, varios estados alemanes y numerosas ciudades italianas [3]. Actualmente con el desarrollo de poder de computo y la disponibilidad de bases de datos empíricos de distintas medidas de interacciones sociales ha sido posible descubrir los patrones de comportamiento e investigar las causas de la desigualdad socio económica. Dos tipos de desigualdad que podemos resaltar son la desigualdad de condiciones, la cual consiste en la distribución desigual del ingreso, la riqueza y en general de los bienes materiales de los que se dispone. La segunda es la desigualdad de oportunidades la cual consiste de una desigualdad de oportunidades de vida, la cual se refleja en el nivel educativo, la salud e incluso el trato que se recibe de parte de la justicia penal. El estudio de la desigualdad es de vital importancia puesto que es responsable de conflictos, guerras, opresión, actividades ilegales, inestabilidad así como también de afectar el crecimiento

económico [5].

Tradicionalmente la desigualdad que se mide es la de ingreso o riqueza en una sociedad, está se pueden medir de distintas formas, sin embargo, las medidas mas populares consisten en índices, medidas absolutas asociadas al nivel de desigualdad y que usualmente son valores en el intervalo $[0, 1]$. Entre estos índices el que destaca por ser ampliamente utilizado, es el índice Gini, el cual esta relacionado con la curva de Lorenz de manera directa.

2.2. Curva de Lorenz

Existen distintas formas de representar gráficamente la distribución de ingreso entre los distintos grupos que componen una sociedad. La forma mas sencilla sería una gráfica de frecuencias sin embargo este tipo de representación conlleva algunas desventajas como la perdida de información a cerca de la cola de la distribución. Ademas del hecho de que al agrupar los datos empíricos en intervalos, estos quedan representados por su marca de clase lo cual implica necesariamente una perdida de información, más aun, si son pocos los intervalos del gráfico dicha perdida es mayor. La manera habitual en que se representa la desigualdad es por medio de la curva de Lorenz la cual fue propuesta en 1905 para mostrar la desigualdad en la distribución de la salud, desde entonces se ha popularizado en los estudios de desigualdad económica. Para construir la curva de Lorenz partamos de los siguiente: supongamos que tenemos N agentes ordenados de menor a mayor respecto a su ingreso $y_1 \leq y_2 \leq y_3 \cdots y_n$. La curva de Lorenz se define como la relación entre las porciones acumuladas de la población $\%P_i$ y las proporciones de ingreso $\%Y_i$ donde estas se definen como:

$$\begin{aligned} P_i &= \frac{n_1 + n_2 + n_3 + \cdots + n_i}{N}, \\ Y_i &= \frac{y_1 + y_2 + y_3 + \cdots + y_i}{Y} \end{aligned} \tag{2.1}$$

el valor P_i representa el porcentaje acumulado de individuos con ingresos menores o iguales que Y_i y Y_i el porcentaje acumulado de ingresos sobre el total correspondiente a este conjunto de individuos.

Definición 2.2.1. *Se denomina curva de Lorenz a la poligonal formada por los puntos:* $(P_i, Y_i), i = 1, 2, 3, \cdots, N$

que se define en un cuadro $[0, 1] \times [0, 1]$.

Si a cada porcentaje de la población le corresponde el mismo porcentaje de ingreso ($P_i = Y_i, \quad \forall i$) se forma una línea de 45 grados como se muestra en la figura 2.1. Esta linea divide en dos partes iguales el cuadrado de lados uno que se forma al graficar las proporciones acumuladas de personas en el eje horizontal (P_i) y de ingreso en el

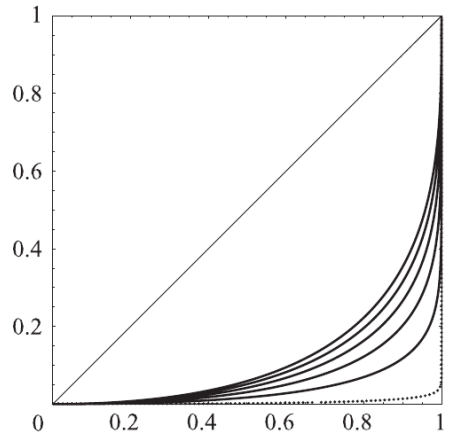


Figura 2.1: Se muestran distintas curvas de Lorenz las cuales corresponden a diferentes distribuciones de ingresos, la linea recta corresponde a la ausencia de desigualdad.

vertical (Y_i).Dicha diagonal corresponde a lo que Lorenz definió como la linea de equidad perfecta y denota por ende ausencia de desigualdad. En el gráfico podemos observar tres elementos: la línea de equidistribución, las curvas correspondientes a las distribuciones empírica formadas a partir de las parejas (Pi, Yi) o curvas de Lorenz, el área entre la linea recta y las curvas de Lorenz se denominada área de concentración. A partir de la relación entre las últimas y la línea de igualdad perfecta, es posible derivar diversos indicadores que se utilizan para evaluar la concentración del ingreso. La curva de Lorenz siempre se ubica por debajo de la diagonal en la medida que los ingresos de los individuos se hayan ordenado en forma creciente y por encima de haberlo hecho de forma contraria. En la medida que la curva de Lorenz se aproxime a la diagonal, observaríamos situaciones de mayor igualdad, mientras que cuando se aleja, la desigualdad se incrementa. El punto (0,0) significa que el 0 % de la población tiene el 0 % del ingreso, mientras que en el extremo opuesto el 100 % de la población concentra todo el ingreso. Si una curva queda totalmente contenida en otra (a excepción de los valores extremos) podemos afirmar, que aquella que se ubica más cerca de la diagonal presenta una distribución más igualitaria, en cuyo caso se dice que domina en el orden de Lorenz.

2.3. Índice de Gini

Gini en 1912 definió su medida de desigualdad de la siguiente forma:

$$CG = \frac{1}{2\mu}\left[\frac{\Sigma_{i=1}^{n}\Sigma_{j=1}^{n}|y_i - y_j|}{n\,(n-1)}\right] = \frac{1}{2\mu}\triangle \tag{2.2}$$

donde $\triangle$ representa la media aritmética de las $n(n-1)$ diferencias absolutas de

las observaciones y 2μ es el valor máximo que asume $\triangle$ cuando todo el ingreso es concentrado por un individuo.

Dos años después, Gini propone una nueva medida que se relaciona con la curva de Lorenz y demuestra que es equivalente a la expresión que había dado. La expresión es la siguiente:

$$CG = 1 - 2f(y) \tag{2.3}$$

donde $f(y)$ representa la curva de Lorenz.

Capítulo 3

Redes

En este capítulo abordamos las definiciones básicas para describir y analizar las redes, la mayoría de estas proviene de la teoría de gráficas, rama de las matemáticas que trata con estas. Dicha rama es un campo bastante amplio que contiene numerosos resultados, sin embargo aquí solo abordamos una pequeña fracciones de estos, centrandonos en los que serán de utilidad para los fines de este trabajo. En este sentido, una tipo de red que tiene importancia para nuestro desarrollo es la red tipo árbol, dicha estructura esta presente en muchos y variados lugares de nuestra vida cotidiana como son: la estructura de archivos de una computadora , las filogenias de los seres vivos y los ríos, solo por citar algunos ejemplos. La definición de árbol de la teoría de gráficas, nos permite darle una definición exacta a nuestras redes jerárquicas. Pero antes de comenzar vale la pena, mencionar algunos puntos sobre el desarrollo que han tenido las redes en la actualidad.

3.1. La ciencia de redes

A medida que nos hemos desarrollado como sociedad es mas evidente nuestra dependencia de sistemas conformados por una gran cantidad de elementos. Los cuales requieren de una cooperación conjunta para funcionar, un ejemplo de estos son los sistemas de comunicación modernos, los cuales están conformados por millones de teléfonos móviles que dependen del funcionamiento de computadoras y satélites para hacer posible que millones de personas puedan realizar llamadas telefónicas diariamente. Este tipo de sistemas no son exclusivos de las creaciones del hombre, el cerebro humano por ejemplo, tiene garantizada su actividad coherente por el funcionamiento de millones de neuronas. A toda esta gama de sistemas se les ha llamado en conjunto sistemas complejos y dada la importancia que estos juegan en nuestra vida diaria, poderlos entender, matematizar, predecir y eventualmente controlar es uno de los grandes retos del siglo XXI. La ciencia a tomado este desafío y una de las formas para enfrentarlo es la "ciencia de redes", de hecho, detrás de cada sistema complejo, hay una intrincada red que encripta las interacciones entre los componentes de estos sistemas. Pongamos algunos ejemplos, las redes que describen las interacciones entre proteínas, genes y procesos

metabólicos en las células. La red de comunicación la cual describe como interactuan los distintos dispositivos de comunicación. La red de interacción de las neuronas en nuestros cerebros. Las redes de comercio que mantienen el intercambio de bienes y servicios que por una parte permiten que disfrutemos de mas comodidad y por otra la propagación de crisis. A pesar de la diversidad en la forma, el tamaño, la edad, la naturaleza y el alcance que caracterizan a las redes reales, la mayoría de las que se han observado en la naturaleza, en la sociedad y la tecnología son impulsadas por principios organizativos comunes. En la referencia PONERREF(24 July 2009 vol 325, issue 5939, pages 357-504 special issue) podemos encontrar algunos ejemplos de sistemas complejos y redes.

La "ciencia de redes" es un campo emergente de estudio de naturaleza principalmente interdisciplinaria, dicho nombre ha generado polémica entre ciertos sectores, puesto que su diferencia con la teoría de gráficas llega a ser sutil. Nosotros consideraremos a la ciencia de redes como un resurgimiento de la teoría de gráficas (con una fuerte base empírica) la cual remonta sus orígenes hasta 1736, cuando el matemático suizo Leonhard Euler resolvió el problema a cerca de cual seria la mejor forma de circunnavegar los puentes de Königsberg, usando una representación de puntos y lineas para darle solución a este problema. Desde sus orígenes la teoría de gráficas a estado muy vinculada a resolver problemas prácticos pero podríamos decir que ha tenido periodos en la que resurge y el último de estos se da a finales de los años 90's. Después de la propuesta de Euler la teoría de gráficas paso los siguientes 200 años solo conocida por un circulo muy reducido. Sin embargo, esta aparece nuevamente en la década de los 50's cuando Paul Erdos (1913-1996) la restablece con sus artículos sobre redes aleatorias. Para la década de 1960 y 1970 esta teoría se utilizó por científicos sociales para modelas redes sociales y estudiar el comportamiento de grupos de humanos. A Stanley Miligram de debemos el introducir la noción de redes de mundo pequeño en la comunidad de las ciencias sociales. Su tercera y actual etapa a través de la cual se ha convertido en una disciplina científica en si misma, se da a finales de los 90's cuando varios científicos de distintas disciplinas comienzan a usar las redes como modelos de fenómenos físicos y biológicos. Por citar algunos nombres de los pioneros en esta disciplina tenemos a : Duncan Watts, Steven Strogatz y Albert-Laslo Barabasi que estimularon el renovado interés en el análisis matemáticos de la redes aplicadas al mundo físico. Watts [7] equipara las redes con escasos vínculos o "sparse" con diámetro pequeño (mundo pequeño) con un número de diversos fenómenos tales como: transiciones de fase en materiales, funcionalidad de los organismos biológicos y el comportamientos de redes eléctricas. Strogaz [6] por su parte estudió el impacto de la estructura de la red en los sistemas adaptativos complejos de la física, así como la explicación de por que los corazones laten en un patrón regular sincronizado en los mamíferos y por que ciertas especies de luciérnagas emiten chirridos rítmicamente al unisono y sin un control centralizado. Barabási y sus estudiantes [1] crearon otra linea de investigación con la propuesta de las redes libres de escala , redes no aleatorias en los que se encuentra "hubs". En una serie de estudios de Internet y de WWW, descubrieron una propiedad emergente de la Internet la cual surge sin planeación central en una estructura, la cual consiste en un pequeño numero de sitios muy populares llamados "hubs" y un gran número de sitios

"impopulares" con pocos enlaces. En lugar de ser aleatoria como en la red de Erdos-Rényi (ER), en Internet la probabilidad de que un sitio tenga k vínculos obedece a una ley de potencias, la cual cae rápidamente para grandes k's. A demás especularon sobre la regla micro que llevaba a este resultado a la cual llamaron vinculación preferencial (preferntial attachment) que dice que la probabilidad de que un sitio tenga un nuevo vinculo es directamente proporcional a la cantidad de vínculos que ya posee. El llamado principio de "los ricos se hacen más ricos".

3.2. Representando las interacciones en los sistemas como redes.

Es gracias a la simplicidad de la teoría gráficas que de manera sencilla se ha podido dar una representación de red a distintos y variados sistemas compuesto de millones de componentes que interactúan entre si. Los componentes de una red (nodos y vínculos) pueden describir distintas unidades del mundo real, como podrían ser, los proveedores de Internet, los generadores de electricidad, los agentes económicos, las especies biológicas de un ecosistema, etc. Los vínculos por otra parte entre los distintos componentes pueden describir un comportamiento global, como el trafico de Internet o el del suministro la electricidad, las tendencia de los mercados, el agotamiento de los recursos naturales, etc. Por ejemplo, para formar la representación en red de un sistema social, requeriríamos la lista de nuestros amigos y la de los amigos de nuestros amigos y de esta forma sucesivamente. De manera intuitiva es claro que la forma de la red y su funcionalidad son cosas que están relacionadas, es decir, si conocemos las propiedades topológicas de una red podríamos determinar leyes que gobiernan al sistema.

3.3. Rasgos que distingue a la ciencia de redes.

No es de sorprender todas las diferencias que se pueden encontrarse entre los componentes de las redes que están detrás de los sistemas complejos. Por ejemplo, en una red metabólica los nodos son moléculas y los vínculos son las reacciones químicas que las gobiernan; en la WWW los nodos son los documentos web y los vínculos son las URL's que son mantenidos por algoritmos computacionales; en las redes sociales los individuos son los nodos y los vínculos pueden ser representados por las relaciones sociales, laborales, de amistad o simplemente el hecho de conocer a alguien. También los procesos que dieron origen a estas son muy variados; en el caso de las redes metabólicas ha sido el proceso de millones de años de evolución, en la WWW la acción colectiva de millones de usuarios. Un descubrimiento clave del estudio actual de las redes es que tanto la topología y la evolución de las redes que emergen de distintos ámbitos como la ciencia, la naturaleza, la sociedad y la tecnología son bastante similares entre sí. que se puede usar un conjunto de herramientas matemáticas para explorar estos sistemas. Esta universalidad es la base que ha guiado el desarrollo de la ciencia de redes actual. Y la forma en que aborda este fenómeno es lo que la distingue. En su libro (NETWORK

SCIENCE CAPITULO 2), Barabási cita las siguientes características de la ciencia en redes, las cuales enlistaremos a continuación para clarificar sus métodos.

- **Naturaleza interdisciplinaria.**Ofrece un lenguaje a traves del cual diferentes disciplinas pueden interactuar entre sí. De hecho, tanto los biólogos celulares y los cientificos de la computación se enfrentan de manera similar a la tarea de caracterizar los vinculos detras de su sistema, a partir de conjuntos de datos incompletos y ruidosos, asi como tambien a tratar de entender la solidez de sus sistemas a fallas o ataques deliberados. A pesar de que cada disciplina tiene sus detalles y desafios tecnicos el caracter comun de los diversos temas han llevado a una gran variedad de ideas y herramientas disponibles para enfrentarlos.

- **Datos empiricos.** Lo que distingue la teoria de graficas de la actual ciencia de redes es su caracter empirico, esto quiere decier que se centra en los datos y su utilidad. De hecho nunca se esta satisfecho con el desarrollo de herramientas matemáticas abstractas para describir una propiedad determinada de la red. Cada herramienta que se desarrolla se pone a prueba en los datos reales y su valor se juzga por los puntos de vistas que ofrece sobre la estructura o la evolución de un sistema.

- **Naturaleza matemática y cuantitativa.** Para poder contribuir a la ciencia de redes es necesario dominar las herramientas matemáticas detras de esta. Esta area toma prestado el formalismo de la teoría de graficas para tratar con las graficas y los conceptos de alatoreidad para buscar principios organizativos universales de la fisica estadistica. Ultimamente el campo se ha beneficiado de conceptos tomados de la ingenieria del control y la teoria de la información, la estadistica y la mineria de datos , lo que ha ayudado a extraer la informacion de conjuntos de datos incompletos y ruidosos.

- **Nauturaleza computacional.** Dado el tamaño de muchas de las redes que se analizan y la excepcional cantidad de datos detrás de estas, se plantean una serie de retos computacionales. Por lo tanto, el campo tiene un carácter computacional fuerte, tomando prestados de forma activa algoritmos para el manejo de las bases de datos y minería de datos. Una serie de herramientas de sofware ayudan a los profesionales con diversas habilidades computacionales a analizar las redes.

3.4. Algunas representaciones de las redes.

Una gráfica o red es una colección de nodos unidos por vínculos. A menudo estos dos conjuntos son llamados V (conjunto de nodos) y E (conjunto de vínculos). La gráfica se indica por $G(V, E)$, donde los nodos pueden ser representados por puntos y los vínculos como lineas entre ellos. No es importante como se dibujen, en última instancia lo más importante es que nodos están conectados y con quien. Se denotara por N a el número de nodos y a L el número de vínculos. Las redes que vamos a utilizar solo tendrán un único vínculo entre cualquier par de nodos. Sin embargo, hay que mencionar

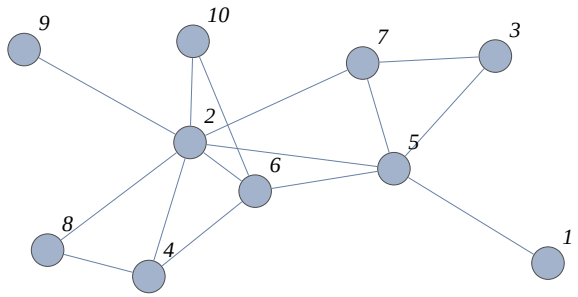


Figura 3.1: Una red simple, no contiene *multivínuculos* ni *autobucles*.

que las redes pueden tener mas de un vínculo entre el mismo par de nodos, este tipo de vínculos recibe el nombre de *multivínculos*. Otro caso, que no usaremos pero que llega a presentarse en las redes es el caso de un vínculo con el mismo nodo, este tipo de enlace es llamado *autobucle*. A partir de las caracteristicas de los enlaces las redes reciben distintos nombres, llamaremos red o gráfica simple a las redes que no tienen ni *multivínculos* ni *autobucles* ver figura 3.1.

Existen distintas formas de representar una red matemáticamente. Por ejemplo si consideramos una red simple no direccionada con N nodos a los cuales los etiquetamos con un numero entero $1, 2, 3, \cdots, N$,estas etiquetas no tiene mayor importancia, salvo que es única para cada nodo y por lo tanto, nos podemos referir aun nodo sin confusión. Una forma usual de representar a la gráfica es por medio de una lista de pares que representan los vínculos entre los nodos que la conforman. Por ejemplo, si llamamos al vínculo entre el nodo i y el nodo j como (i, j) entonces la gráfica de la figura 3.1 queda representada por el valor de N y la lista de los vínculos, que para este caso sería:

$$(1,5),(5,3),(5,7),(5,2),(5,6),(3,7),(7,2),(6,10),$$
$$(6,2),(6,4),(2,10),(2,9),(2,4),(2,8),(8,4)$$

este tipo de representación recibe el nombre de lista de vínculos.

3.4.1. Matriz de adyacente.

Otra forma de representar una red es por medio de la matriz de adyacencia $\mathbf{A}$, donde los elementos de esta se definen como:

$$A_{ij} = \begin{cases} 1 & si\, existe\, un\, vinculo\, entre\, i\, y\, j \\ 0 & en\, otro\, caso \end{cases} \tag{3.1}$$

.

La matriz de adyacencia **A**, para nuestra red quedaría como :

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{3.2}$$

Como podemos observar los elementos de la diagonal de la matriz son cero, esto se debe a que nuestra red no contiene *autobucles*. Otra característica de esta matriz es su simetría, esto es debido a que es una red no direccionada, es decir, el nodo i esta relacionado con el nodo j, pero también el nodo j esta relacionado con el nodo i.

3.4.2. Redes pesadas.

En algunas situaciones es útil asignar a los vínculos un valor que represente una propiedad de la red, como por ejemplo, a la red de suministro de agua le podemos asociar el flujo que hay de un nodo a otro. Otro ejemplo, que se puede ver en artículos de redes alimenticias, es cuando a cada vínculo se le asigna el valor de la energía que fluye de la presa al depredador. A las redes a las cuales se les asocia un valor a los vínculos se les llama *redes pesada*. Este tipo de red también puede ser representada por una matriz adyacente, pero en estas los valores de las entradas son distintos de cero ver en la matriz 3.3.

$$\mathbf{A} = \begin{pmatrix} 0 & 3 & 0 & 0 & 0.5 \\ 3 & 0 & 1.2 & 0 & 2 \\ 0 & 1.2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0.5 & 2 & 0 & 0 & \end{pmatrix} \tag{3.3}$$

nuevamente podemos observar en la matriz que esta es simétrica y que no tiene autobucles, aunque en principio no habría problema en considerar este tipo de redes.

3.4.3. Redes direccionadas

Una red direccionada o gráfica direccionada, es una red en la cual cada vínculo tiene una dirección que puede ser representada por una flecha que va de un nodo a otro, como la que se muestra en la figura 3.2.

Hay numerosos ejemplos de redes direccionadas, por ejemplo la WWW donde los hipervínculos te dirigen de una pagina web a otra, pero no necesariamente en esta última existe un vínculo que te dirija a la pagina inicial. Twitter es otro ejemplo de

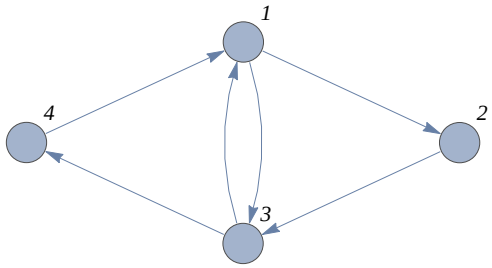


Figura 3.2: Una red direccinada, las flechas representan la dirección.

red direccionada. Este tipo de redes también pueden ser representadas por una matriz adyacente y sus elementos estarían dados por:

$$A_{ij} = \begin{cases} 1 & si\, existe\, un\, vinculo\, de\, j\, a\, i \\ 0 & en\, otro\, caso \end{cases} \tag{3.4}$$

.

Como podemos ver la dirección del vínculo va del segundo índice al primero. La matriz adyacente para la figura 3.2 sería:

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix} \tag{3.5}$$

En general la matriz adyacente de una red direccionada es asimétrica.

3.5. Medidas de las redes.

Una medida importante para cada uno de los nodos es su *grado*, este representa el número de vínculos que tiene con otros nodos. Ejemplos del grado en redes reales pueden ser el número de contactos que se tienen en tu correo electronico o el telefono movil o el numero de citas que tiene un articulo. Usualmente el grado de un nodo $i'esimo$ es denota por la letra k_i. Este puede escribirse en términos de la matriz de adyacencia de la siguiente forma:

$$k_i = \sum_{i}^{N} A_{ij}. \tag{3.6}$$

En una red no direccionada el número de vínculos L se puede representar como la suma de los grados de los nodos:

$$L = \frac{1}{2} \sum_{i}^{N} k_i \tag{3.7}$$

el factor 1/2 en esta expresión, corrige el hecho de que el grado de cada nodo es sumado dos veces. El grado promedio lo podemos expresar en términos de los vínculos de la siguiente forma:

$$\langle k \rangle \equiv \frac{1}{N} \sum_{i}^{N} k_i = \frac{2L}{N} \tag{3.8}$$

Para el caso de redes direccionadas podemos distinguir entre los dos tipos de vínculos, los que entran al nodo y los que salen. Entonces, llamamos k_i^{in} al número de vínculos que apuntan hacia el nodo $i'esimo$ y k_i^{out} el número de vínculos que apuntan desde el $i'esimo$ nodos, el grado total sería de nodo $i'esimo$ sería,

$$k_i = k_i^{in} + k_i^{out}. \tag{3.9}$$

Ahora, el total de vínculos para el caso de redes direccionadas es:

$$L = \sum_{i}^{N} k_i^{in} = \sum_{i}^{N} k_i^{out} \tag{3.10}$$

que como podemos notar, el factor 1/2 no aparece como en la ecuación 3.7, dado que los grados de entrada y salida se cuentan por separado.

El grado promedio para una red direccionada :

$$\langle k^{in} \rangle = \frac{1}{N} \sum_{i}^{N} k_i^{in} = \langle k^{out} \rangle = \frac{1}{N} \sum_{i}^{N} k_i^{out} = \frac{L}{N} \tag{3.11}$$

La distribución del grado p_k, da la probabilidad de que un nodo elegido aleatoriamente en una red tenga grado k. De la condición de normalización para la probabilidad p_k, tenemos que :

$$p_k = \frac{N_k}{N}, \tag{3.12}$$

donde N_k es el número de nodos de grado k. De esta forma si conocemos la distribución de probabilidad del grado podemos obtener el número de nodos de grado k a partir de la distribución como $N_k = Np_k$.

Ahora, el grado promedio $\langle k \rangle$ de una red se puede escribir en términos de la distribución:

$$\langle k \rangle = \sum_{i}^{\infty} k p_k. \tag{3.13}$$

La forma precisa de la la distribución de probabilidad p_k determina muchos de los fenómenos de las redes.

Otra caracteristica de las redes simples a destacar es el número máximo de vínculos:

$$\binom{N}{2} = \frac{1}{2}N\left(N-1\right). \tag{3.14}$$

La densidad ρ de una red es la fracción de vínculos presentes entre estos últimos (ecuación 3.14):

$$\rho = \frac{L}{\binom{N}{2}} = \frac{2L}{N(N-1)} = \frac{\langle k \rangle}{N-1}, \tag{3.15}$$

la densidad se encuentra en el rango $0 \leq \rho \leq 1$. Una red que tiende a una constante a medida que $n \longrightarrow \infty$ decimos que es densa y una red en la que $\rho \longrightarrow 0$ a medida que $n \longrightarrow \infty$ se dice que es "sparse".

3.6. Distancia en redes

La distancia física que usualmente se utilizamos en para describir objetos, no es útil para el caso de redes, en estas la definición de distancia se sustituye por otro concepto al cual se le llama *longitud de camino*. Un *camino* es una ruta que va a lo largo de los vínculos de una red, su longitud se representa por el numero de vínculos que contiene ese camino. A continuación damos algunas propiedades importantes de este.

Camino mas corto (o camino geodésico) entre los nodos i y j es el camino con el menor número de vínculos posibles entre dos nodos. El *camino mas corto* a menudo se le llama la distancia entre los nodo i y j y se representa por d_ij o simplemente d. Es usual que en las redes se lleguen a encontrar varios caminos de la misma d entre algún par de nodos pero no lo que no deben contener estos son bucles o intersecciones entre ellos.

Para una red no direccionada $d_ij = d_ji$, es decir la distancia entre el nodo i y el nodo j es la misma. Esto no se cumple en general para redes direccionadas en general en estas $d_ij \neq d_ji$. De hecho, la existencia de un camino del nodo i al nodo j no es garantía de que exista uno del nodo j al nodo i.

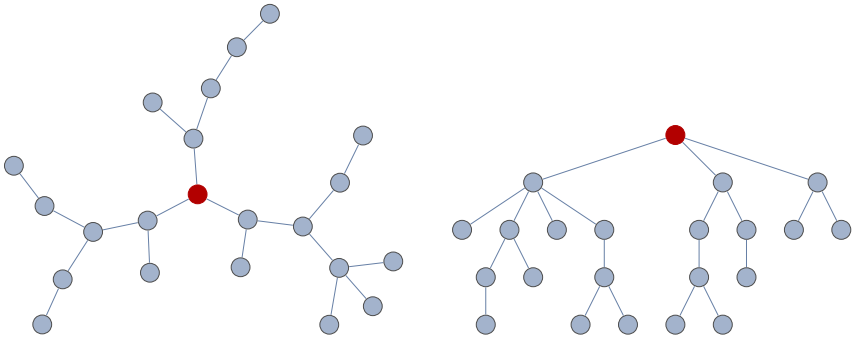
Diametro de una red. El diámetro de una red se representa por d_max, el cual es el máximo de los caminos cortos. Es decir, la distancia mas grande entre cualquier par de nodos.

Longitud de camino promedio. Se representa por $\langle d \rangle$, es la distancia promedio entre todos los pares de nodos en una red. Para una red direccionada de N nodos $\langle d \rangle$ esta dada por:

$$\langle d \rangle = \frac{1}{N(N-1)} \sum_{i,j=1,N} \mathbf{d}_{i,j} \tag{3.16}$$

3.7. Arboles

Los arboles son redes conectadas, pueden ser direccionadas o no direccionadas que no contienen bucles. Cuando decimos conectadas nos referimos a que a cada nodo de



Cuadro 3.1: Mostramos dos árboles en el de la derecha se ha elegido a uno de los nodos como raíz y se acomoda la estructura de tal forma que este se encuentre en la parte mas alta. Este tipo de representación llega a ser útil si en el sistema a representar, uno de los nodos tiene propiedades particulares. Como el jefe en un organigrama, el nacimiento de un río, etc.

la red se puede llegar desde otro por medio de un camino. Este tipo de redes también pueden consistir de varias partes desconectadas una de la otra, en es caso se les llama bosques.

Los arboles puede ser representado de manera enraizada, es decir se elige un nodo para para aparecer en la parte mas alta de la estructura de ramificación. Los nodos que se encuentran en la parte mas baja de esta estructura y que solo tiene un vínculo se les llama hojas mientras que al resto nodos o nodos internos. Elegir un nodo como el nodo raíz no tiene importancia para la definición de árbol, sin embargo para algunas aplicaciones es importante hacer esta distinción, este es el caso de los árboles taxonomicos, donde el nodo raíz es la especie madre o en un río donde la fuente sería el nodo raíz.

Una de las propiedades mas importantes de los arboles es que puesto que estos no tienen bucles cerrados, existe exactamente un camino entre dos pares de nodos. Otra muy útil es que un árbol de N nodos siempre tiene exactamente $N-1$ vínculos. De hecho, el regreso de esta afirmación también es verdadera. Para una red conectada de N nodos y $N-1$ vínculos es un árbol.

Capítulo 4

Desigualdad en redes jerárquicas

Como hemos visto, uno de los indices mas utilizados (tanto por organismos internacionales y gobiernos como por organizaciones sociales) para la medición de la desigualdad, por su fácil obtención, como por su relación con la curva de Lorenz (la cual permite una representación bastante clara e ilustrativa), es el indice de Gini. Sin embargo, a diferencia de dicho indice, el cual da un valor en términos de la distribución de una variable sobre los elementos de un conjunto, a nosotros nos interesa proponer un indice de desigualdad en términos de las posiciones que ocupan estos elementos en una estructura de tipo jerárquica, es decir, la forma en que están relacionados los elementos del conjunto, como la posición que ocupan dentro de la estructura tendrán un papel prioritario para la definición de nuestro indice de desigualdad. En este punto surgen las siguientes preguntas ¿qué tan jerárquica es una estructura?, ¿de qué forma la topologia de la red y la posición de los elementos pueden tomarse en cuenta para la definición de desigualdad? y ¿como normalizamos para que los valores de nuestro indice tome valores en el intervalo $[0, 1]$ para las distintas estructuras jerarquicas?. En este capitulo desarrollaremos estas preguntas, asi como tambien calcularemos el indice de desigualdad propuesto para estructuras jerarquicas conocidas.

4.1. Distancia en estructuras jerárquicas

Cuando nos referimos a redes jerárquicas nos referimos a *arboles con raíz*, dichas estructuras estan presentes en diversos campos del conocimiento, desde la Teoria de graficas en Matemáticas, pasando por las estructuras de datos de las Ciencias de la Computación, hasta los jerárquigramas en las Ciencias Sociales. Sin embargo, las redes que abordaremos en este trabajo, son redes jerárquicas que han adquirido su forma (topologia) y numero de nodos como resultado de procesos evolutivos constantes y de los cuales contamos con información en un punto en el tiempo, con la cual somos capaces de construir la red jerárquica, ejemplos de estás son, los arboles genealogicos de los lenguajes, las organizaciones sociales, privadas o estatales e incluso con algunas particularidades las redes alimenticias en un ecosistema.

Para comenzar es necesario establecer que entendemos por distancia en las redes

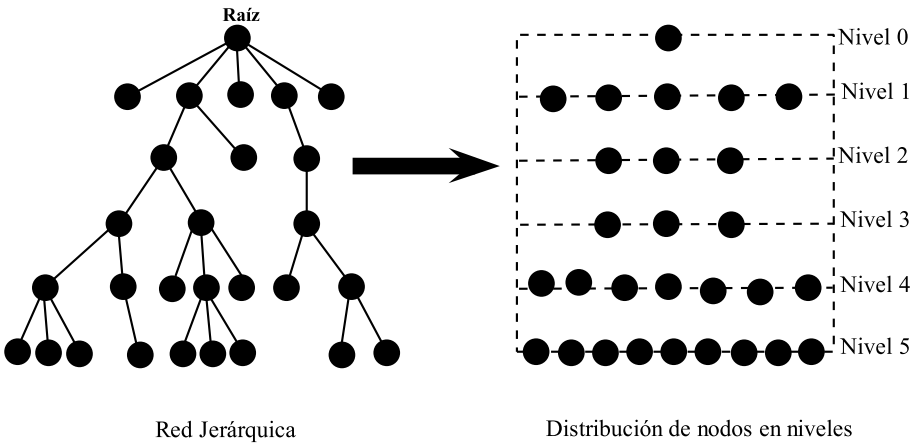


Figura 4.1: A partir de la red jerárquica podemos generar una distribución en niveles, donde el nodo raíz es el único que se encuentra en el $Nivel0$.

jerárquicas, para este propósito, utilizamos la definición de distancia entre nodos para arboles con raíz, que es el número de vínculos que son necesarios recorrer desde el nodo raíz (el cual es designado como tal, por las propiedades que presenta dentro del sistema) hasta cualquier nodo. Como vimos en el Capitulo **??** esta distancia es única para cada nodo del árbol con raíz. De esta forma a cada nodo le podemos asociar una distancia y nodos con la misma distancia decimos que se encuentran en el mismo nivel o estrato. Por definición el nodo raíz es el único nodo que esta a distancia 0 y por lo tanto se encuentra en el $Nivel\,0$, nodos que tienen una distancia 1 decimos que se encuentran en el $Nivel\,1$ y así sucesivamente para todos los nodos de la red. Entonces podemos representar una red jerárquica de una forma mas simplificada por medio de una distribución de nodos en niveles, como se puede ver en la figura 4.1. La distribución de los nodos en los distintos niveles depende del árbol en particular y es posible encontrar varios arboles con la misma distribución de nodos. Para nuestro análisis los vínculos solo tendrán importancia en el sentido que estos determinan una distribución de los nodos en distintos niveles o estratos, a partir de la cual deseamos construir un índice que refleje esta desigualdad posicional.

4.1.1. Diferencia promedio de la distribución de N nodos $\langle d_N \rangle$

Una pregunta que surge casi inmediatamente después de haber definido la distancia entre el nodo raíz y cualquiera de los nodos, es ¿cuál es la distancia entre dos nodos cualquiera?, en este punto, reflexionamos de la siguiente forma: dado que la propiedad que deseamos que refleje nuestra medida a construir es que tan separados están los nodos unos de otros, respecto a la posición que ocupan en una red jerárquica, es natural establecer que la distancia entre dos nodos es la diferencia entre los niveles que estos ocupan, es decir, la distancia entre en nodo i y el nodo j se define como la diferencia entre los niveles que ocupan $|n_i - n_j|$. Entonces una manera de asociar una medida

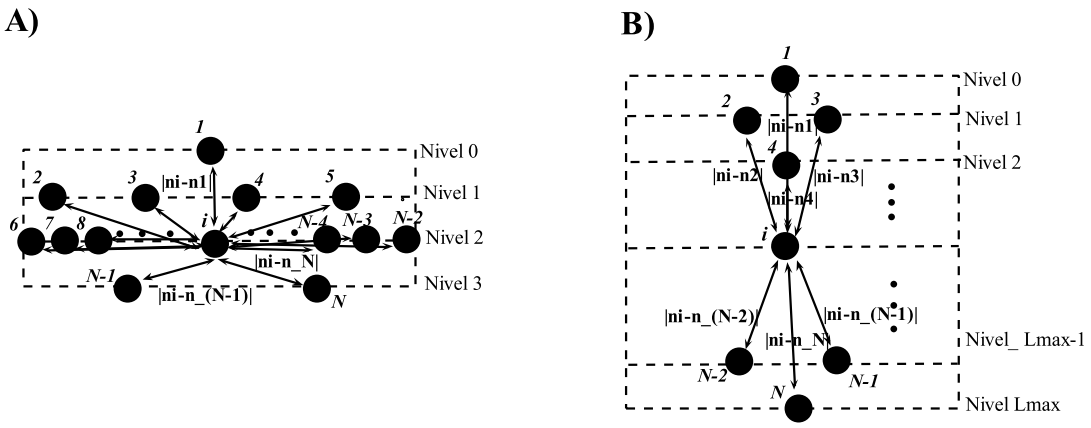


Figura 4.2: **A)**Observamos una distribución de N nodos en tres niveles, las diferencias en las distancias, las cuales están representadas como flechas para el caso del nodo i, varían desde 0 (nodos en el mismo nivel) hasta 3 (la distancia del nodo raíz al nodo N. **B)** Distribución de N nodos en L_{max} niveles, donde $3 \ll L_{max} \leq N-1$, las diferencias entre distancias varían dese 0 hasta L_{max}. Puesto que la sumatoria de las diferencias de las distancias es mayor para el inciso **B)** que para el **A)** tenemos entonces que $\langle d_N \rangle_B > \langle d_N \rangle_A$.

global a la distribución de los nodos en los distintos niveles, es por medio del promedio de las diferencias entre sus distancias, que queda definido de la siguiente forma:

$$\langle d_N \rangle = \frac{\sum_{i \neq j=1}^{n} |n_i - n_j|}{2N(N-1)}, \tag{4.1}$$

donde, n_i, como mencionamos arriba, es el nivel en que se encuntra el nodo *i'esimo* y N es el total de nodos que forman la red jerárquica. Esta medida nos permite asociar a cada red un valor, que nos da información acerca de que tan alejados están los nodos unos de otros en una red jerárquica. De manera empírica sabemos que una red que tiene la mayoria de sus nodos entorno a un mismo nivel tendrá un valor cercano a 0, mientras que redes con un numero mayor de niveles y con distribuciones mas homogéneas de nodos, se alejaran de este valor, ver figura 4.1.1.

Cuando se construye un índice es deseable que este tome valores en el intervalo $[0, 1]$, con este fin es necesario normalizar nuestro valor promedio. La manera más sencilla por la que optamos en un principio para normalizar el promedio, fue usar el máximo de $|n_i - n_j|, \forall i \neq j$. De esta forma nuestra expresión para la desigualdad (I) quedaría:

$$I = \frac{1}{L} \frac{\sum_{i \neq j=1}^{n} |n_i - n_j|}{2N(N-1)}, \tag{4.2}$$

donde L es el máximo de las diferencias $(|n_i - n_j|, \forall i \neq j)$ en la distribución en niveles.

Sin embargo esta última ecuación nos genero distintos inconvenientes. El primero de ellos es cuando tenemos una distribución en niveles de N nodos; donde el $Nivel\,0$ es ocupado por el nodo raíz y los $N-1$ nodos restantes ocupan el $Nivel\,1$. En este caso el valor del I no es cero, sino un valor muy cercanos a cero ($I \approx 0$). Por otra parte, no es claro, que tipo de estructura jerárquica es necesaria para generar una distribución en niveles con un índice de desigualdad igual a 1. Un $I = 1$ obliga a que el valor promedio de las diferencias de las distancias sean igual que el valor máximo de estas, lo cual es imposible para una distribución en niveles que proviene de una red jerárquica. Lo anterior nos lleva a buscar una mejor forma de normalizar $\langle d_N \rangle$ que se apegue a la idea de que distribuciones con nodos cercanos tengan un índice cercano a 0 y distribuciones con nodos alejados se acerquen a valores de 1.

4.2. Valores extremos en la distribución de nodos

A partir de la ecuación (4.1) y considerando un numero de nodos N fijo, podemos observar que el valor de $\langle d_N \rangle$ sólo depende de la sumatoria de las diferencias entre los niveles que ocupan los nodos de una red jerárquica. Puesto que deseamos que el índice que estamos construyendo tome valores en el intervalo $[1, 0]$ es necesario determinar para cuales distribuciones dicha sumatoria toma el valor máximo y mínimo y cuanto vale $\langle d_N \rangle$ en estos casos.

4.2.1. Valor mínimo de $\langle d_N \rangle$

Para que la sumatoria de diferencias entre niveles tome el menor valor posible, todos los nodos deben estar en un mismo $Nivel$, de esta forma las diferencias serían cero, sin embargo, este caso no es posible, puesto que no existe una una red jerárquica que pueda generar tal distribución. Las condiciones que establece una red jerárquica a la distribución en $Niveles$, como habíamos visto, es que el el primer nivel, el cual llamamos $Nivel\,0$, sólo es ocupado por un único nodo, el nodo raíz y la segunda condición, que entre dos $Niveles$ no puede existir un $Nivel$ vacío. Entonces, para el caso de una red de N nodos la distribución en $Niveles$ que más se aproxima a que todos los nodos ocupen el mismo $Nivel$, es cuando $N-1$ nodos ocupan el $Nivel\,1$ y el $Nivel\,0$ es ocupado por el nodo raíz. Esta distribución corresponde a una red donde hay un nodo raíz y los $N-1$ nodos restantes son $hojas$, como podemos ver en la Figura 4.2.1

Bastaría con colocar uno de los $N-1$ nodos en el siguiente nivel ($Nivel\,2$) para que la sumatoria de diferencias fuese mayor que la anterior. Ahora determinemos el valor de $\langle d_N \rangle$ para este caso:

$$\begin{aligned}\frac{\sum_{i\neq j=1}^{N} |n_i - n_j|}{2N(N-1)} &= \frac{1}{2N(N-1)} \sum_{i\neq j=1}^{N} |n_i - n_j| \\ &= \frac{1}{2N(N-1)}\end{aligned}$$

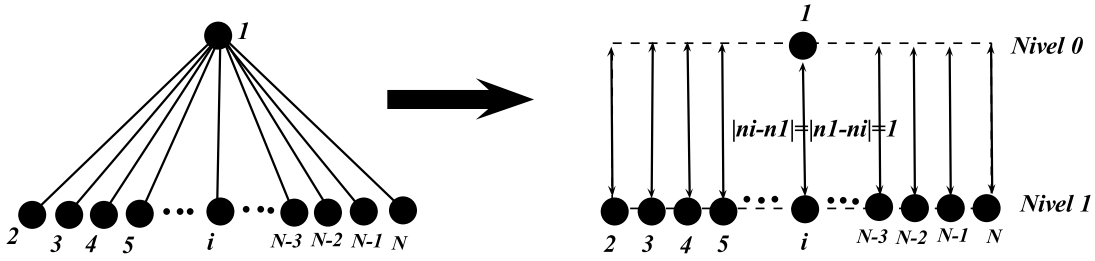


Figura 4.3: Red y distribución de niveles para el caso mínimo

$$\left[\begin{array}{l} \quad |n_1-n_2|+|n_1-n_3|+\ldots+|n_1-n_n| \\ +|n_2-n_1|+|n_2-n_3|+\ldots+|n_2-n_n| \\ +\ddots \\ +|n_i-n_1|+|n_i-n_2|+\ldots+|n_i-n_n| \\ +\ddots \\ +|n_{n-1}-n_1|+|n_{n-1}-n_2|+\ldots+|n_{n-1}-n_n| \\ +|n_n-n_1|+|n_n-n_2|+\ldots+|n_n-n_{n-1}| \end{array}\right]$$

$$= \frac{1}{2N(N-1)}$$

$$\left[\begin{array}{l} \quad 1+1+\ldots+1 \\ +1+0+\ldots+0 \\ +\ddots \\ +1+0+\ldots+0 \\ +\ddots \\ +1+0+\ldots+0 \\ +1+0+\ldots+0 \end{array}\right]$$

Puesto que tenemos $N-1$ sumandos por renglón y N renglones tenemos:

$$\frac{1}{2N\,(N-1)}\sum_{i\neq j=1}^{N}|n_i-n_j| = \frac{1}{\cancel{2}N\cancel{(N-1)}}\cancel{2(N-1)}$$

$$= \frac{1}{N}.$$

$$\therefore \qquad \langle d_{N_{min}}\rangle = \frac{1}{N} \qquad\qquad \forall \quad N \geq 3. \tag{4.3}$$

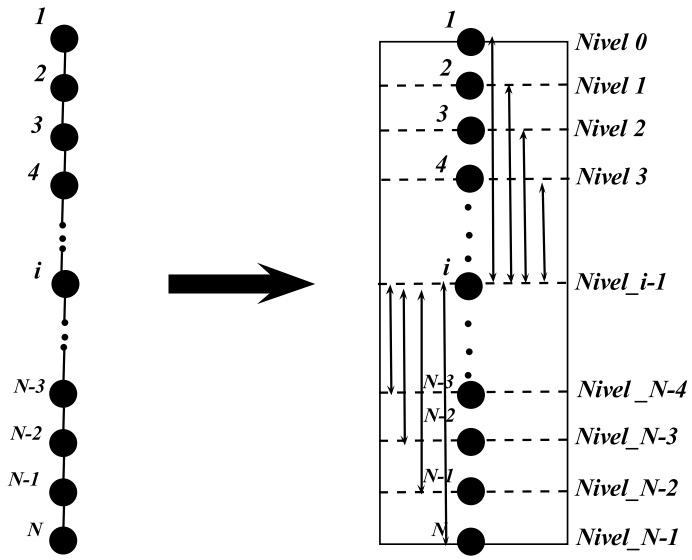


Figura 4.4: Red y distribución de niveles para el caso máximo

4.2.2. Valor máximo de $\langle d_N \rangle$

Ahora, para que la sumatoria de las diferencias entre los niveles que ocupan los nodos tome el mayor valor posible, es necesario, que estos se encuentren lo mas separados entre sí que lo permitan las restricciones de una red jerárquica, esta distribución corresponde al caso en que cada uno de los N nodos ocupen un nivel distinto. Dicha distribución en niveles corresponde a una cadena, es decir, una red jerárquica en la cual el nodo raíz tiene un solo nodo hijo y este a su vez tiene nuevamente un solo nodo hijo y de esta forma sucesivamente hasta completar los $N-1$ nodos siendo el nodo N una hoja, como podemos ver en la Figura 4.4. El argumento para establecer que esta es la distribución cuya sumatoria de diferencias el la máxima posible, es suponer que dos nodos ocupan el mismo nivel, la sumatoria de ésta seria menor que si ocupara un nivel distinto donde él fuera único, puesto que la diferencia entre los dos nodos que ocupan el mismo nivel no aportaría nada a la sumatoria. Para este caso, comenzaremos por calcular primero la sumatoria de diferencias y posteriormente dividir el resultado para obtener el promedio $\langle d_N \rangle$.

Entonces determinamos:

$$\sum_{i\neq j=1}^{N} |n_i - n_j| =$$

$$|n_1 - n_2| + |n_1 - n_3| + |n_1 - n_4| + \ldots + |n_1 - n_N| +$$

$$|n_2 - n_1| + |n_2 - n_3| + |n_2 - n_4| + \ldots + |n_2 - n_N| +$$

$$|n_3 - n_1| + |n_3 - n_2| + |n_3 - n_4| + \ldots + |n_3 - n_N| +$$

$$\ddots$$

$$|n_N - n_1| + |n_N - n_2| + |n_N - n_3| + \ldots + |n_N - n_{N-1}|$$

$$=$$

$$\begin{array}{lllllllllll}
1 & + & 2 & + & 3 & + \ldots + & N-2 & + & N-1 & + \\
1 & + & 1 & + & 2 & + \ldots + & N-1 & + (N-1) - 1 + \\
2 & + & 1 & + & 1 & + \ldots + (N-1) - 1 + (N-1) - 2 + \\
\multicolumn{11}{c}{\ddots} \\
i & + \ldots + & 1 & + & 1 & + \ldots + (N-i) - i + \\
\multicolumn{11}{c}{\ddots} \\
N-2 + \ldots + & 3 & + & 2 & + & 1 & + (N-1) - (N-2) + \\
N-1 + \ldots + & 4 & + & 3 & + & 2 & + & 1
\end{array}$$

,

lo cual lo podemos reducir a

$$=2\left[1+(1+2)+(1+2+3)+\ldots+(1+2+3+\ldots+(N-1))\right]$$

$$=2\left[1(N-1)+2(N-2)+3(N-3)+\ldots+(N-1)(N-(N-1))\right]$$

$$=2\left[N-1+2N-4+3N-9+\ldots+(N-1)N-(N-1)^2\right]$$

$$=2\left[N\sum_{i=1}^{N-1} i-\sum_{i=1}^{N-1} i^2\right]$$

$$=2\left[N\frac{(N-1)N}{2}-\frac{(N-1)(N-1+1)(2(N-1)+1)}{6}\right]$$

$$=2\left[\frac{(N-1)N^2}{2}-\frac{(N-1)(N)(2N-1)}{6}\right]$$

$$=2\left[\frac{3N^3-3N^2-(N^2-N)(2N-1)}{6}\right]$$

$$=2\left[\frac{N^3-N}{6}\right]$$

$$=\frac{N(N^2-1)}{3}.$$

Entonces la suma de diferencias entre niveles para una cadena de N nodos es:

$$\sum_{i\neq j=1}^{N} |n_i-n_j|=\frac{N(N^2-1)}{3}. \tag{4.4}$$

Finalmente dividimos este ultimo resultado para obtener el $\langle d_N\rangle$:

$$\frac{\sum_{i\neq j=1}^{N}|n_i - n_j|}{2N(N-1)} = \frac{\frac{N(N^2-1)}{3}}{2N(N-1)}$$

$$= \frac{N(N-1)(N+1)}{6N(N-1)}$$

$$= \frac{N+1}{6},$$

entonces:

$$\langle d_{N_{max}} \rangle = \frac{N+1}{6} \tag{4.5}$$

4.3. Índice de desigualdad en redes jerárquicas

Finalmente obtendremos la expresión del índice de desigualdad (Id) para redes jerárquicas que proponemos. Como todo buen índice deseamos que los valores que tome se encuentren en el intervalo $[0, 1]$.

Hasta ahora hemos determinado el valor $\langle d_N \rangle$ (promedio de las diferencias entre los niveles de N nodos) para los casos extremos, ecuación 4.3 y 4.5, para el mínimo y el máximo de sus valores respectivamente. Estos valores nos serán útiles para obtener el Id en el intervalo $[0, 1]$ siendo 0 el valor para la mínima desigualdad y 1 para la máxima. Supongamos que tenemos una red jerárquica de N nodos, para la cual calculamos su valor de $\langle d_N \rangle$, dicho valor se encuentra entre los valores:

$$\frac{1}{N} \leq \langle d_N \rangle \leq \frac{N+1}{6}. \tag{4.6}$$

Nosotros deseamos transformar este intervalo en el $[0, 1]$ como podemos ver en la Fig. Para esto hacemos lo siguiente :

$$\begin{aligned}
\frac{1}{N} \leq \langle d_N \rangle \leq \frac{N+1}{6} \\
\frac{1}{N} - \frac{1}{N} \leq \langle d_N \rangle \frac{1}{N} \leq \frac{N+1}{6}\frac{1}{N} \\
0 \leq \langle d_N \rangle \frac{1}{N} \leq \frac{N(N+1)-6}{6N} \\
0 \leq \frac{\langle d_N \rangle \frac{1}{N}}{\frac{N(N+1)-6}{6N}} \leq 1 \qquad \forall N > 2.
\end{aligned} \tag{4.7}$$

El termino central de la desigualdad, es el que definiremos como índice de desigualdad para redes jerárquicas Id:

$$Id = \frac{\langle d_N \rangle \frac{1}{N}}{\frac{N(N+1)-6}{6N}}. \tag{4.8}$$

Sustituyendo el valor de $\langle d_N \rangle$ (ecuación 4.1) en la ecuación 4.8 tenemos:

$$\begin{aligned}
Id =& \frac{6N}{N(N+1)-6}\left[\frac{\sum_{i\neq j=1}^{N}|n_i - n_j|}{2N(N-1)} - \frac{1}{N}\right] \\
=& \frac{6N}{N(N+1)-6}\left[\frac{\sum_{i\neq j=1}^{N}|n_i - n_j| - 2(N-1)}{2N(N-1)}\right] \\
=& \frac{6N}{(N^2+N-6)2N(N-1)}\left[\sum_{i\neq j=1}^{N}|n_i - n_j| - 2(N-1)\right] \\
=& \frac{3}{(N+3)(N-2)(N-1)}\left[\sum_{i\neq j=1}^{N}|n_i - n_j| - 2(N-1)\right]
\end{aligned} \tag{4.9}$$

$$\therefore \qquad Id = \frac{3}{(N+3)(N-2)(N-1)} \left[\sum_{i \neq j=1}^{N} |n_i - n_j| - 2(N-1) \right] \qquad \forall N > 2. \tag{4.10}$$

4.4. Representación de distribución en niveles como $N'ada$

Como mencionamos, cada red jerárquica determina una distribución de nodos en niveles, dicha distribución puede ser representada como una $N'ada$ de valores, donde cada entrada corresponde a un nodo y el valor que está toma, al nivel que ocupa el nodo, por ejemplo: $(0, 1, 2, 3)$ representa a cuatro nodos que están distribuidos en los niveles $1, 2, 3\, y\, 4$ respectivamente. Está interpretación también implica que el orden del arreglo no tiene importancia para nosotros, es decir, $(0, 1, 2, 3)$ significa lo mismo que $(0, 3, 2, 1)$ o cualquiera de sus posibles permutaciones. Para fines ilustrativos, omitamos por un momento el hecho de que el $Nivel\, 0$ solo puede ser ocupado por un nodo; restricción que imponen las redes jerárquicas a las distribuciones en niveles. También pedimos que para que un nodo ocupe un $Nivel$ es necesario que al menos exista otro que ocupe el $Nivel$ anterior (no hay niveles vacíos), casos como $(0, 1, 2, 4)$ o $(0, 2, 3, 5)$ no son posibles. Tomando en cuenta las reglas anteriores, los casos para distribuir 4 nodos son los siguientes:

$$\begin{aligned} &N = 4 \\ &(0,0,0,0),(0,0,0,1),(0,0,1,1) \\ &(0,0,1,2),(0,1,1,1),(0,1,1,2) \\ &(0,1,2,2),(0,1,2,3), \end{aligned} \tag{4.11}$$

el numero de los arreglos para $N = 4$ nodos fue 8 que se puede escribir como 2^3. Esperaríamos que en general el numero de posibles distribuciones para N nodos sea:

$$2^{N-1}, \tag{4.12}$$

para poder obtener este resultado, es necesario considerar los siguientes puntos en la construcción de las distribuciones para un numero fijo N de nodos:

- Supongamos que tenemos N nodos a distribuir en $Niveles$.
- El primer nodo solo puede ocupar el $Nivel$ 0.
- El segundo nodo a colocar tiene dos posibilidades, colocarse en el $Nivel$ 0 o en el $Nivel$ 1.
- El tercer nodo tiene nuevamente dos posibilidades, pero esta dependen de donde se coloco el nodo anterior, si el nodo anterior se coloco en el $Nivel$ 0 entonces el

tercer nodo puede estar en le nivel $Nivel$ 0 o en el $Nivel$ 1, pero si el segundo nodo ocupa el $Nivel$ 1 entonces el tercer nodo tiene la posibilidad de colocarse en el $Nivel$ 1 o en el $Nivel$ 2.

- De esta forma continuamos hasta terminar de colocar N nodos.

La idea de construir la distribución de esta forma refleja una característica de los arboles, un nodo no puede estar en el nivel 2 sin que antes no exista un nodo en el nivel 1 al cual este vinculado. El árbol de decisión para 5 nodos se puede ver en la Figura 4.5

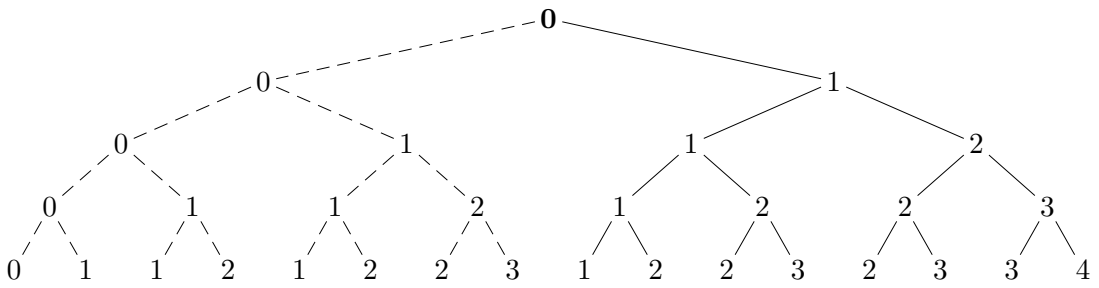


Figura 4.5: Arbol de decision para N=5. El árbol completo representa el caso en que podemos tener mas de un nodo en el *nivel 0*. Si omitimos el subárbol puntueado tendremos el caso en que solo puede haber un nodo en el *Nivel cero* que corresponde a arboles con raíz.

Cada camino de este árbol de decisiones es una posible distribución de $N = 5$ nodos, por ejemplo, el camino $(0, 1, 2, 3, 4))$ corresponde al caso en que solo hay un nodo en cada nivel. Si regresamos a las distribuciones estrictamente de arboles, es decir, distribuciones que permiten solo un nodo en el *nivel* 0, que corresponde al caso en que en la figura 4.5 omitimos el subárbol punteado, tendríamos que cada camino posible correspondería a arreglos que comienzan con un 0. Por ejemplo, a partir del árbol de decisiones construimos las distribuciones posibles para una red jerárquica de $N = 5$ nodos, ver Figura 4.6, como se puede apreciar todas las posibles distribuciones están representadas en el ultimo nivel junto a su $N'ada$ correspondiente del árbol de decisiones.

Regresando a la ecuación 4.12. Del árbol binario de la Figura 4.5 establecimos que el número de posibles distribuciones para N nodos, es igual, al número de caminos que van desde el nodo raíz del árbol de decisiones a cada una de sus hojas en el $Nivel$ $N-1$, como sabemos el número de hojas para un árbol binario crece como potencias de dos, por lo tanto, el número de posibles distribuciones (sin la restricción de que solo un nodo ocupe el $Nivel$ 0) es 2^{N-1}. Ahora para el caso de las redes jerárquicas, podemos establecer lo siguiente: *el número de posibles distribuciones de redes jerárquicas con N nodos es exactamente la mitad de las distribuciones cuando no tiene la restricción de*

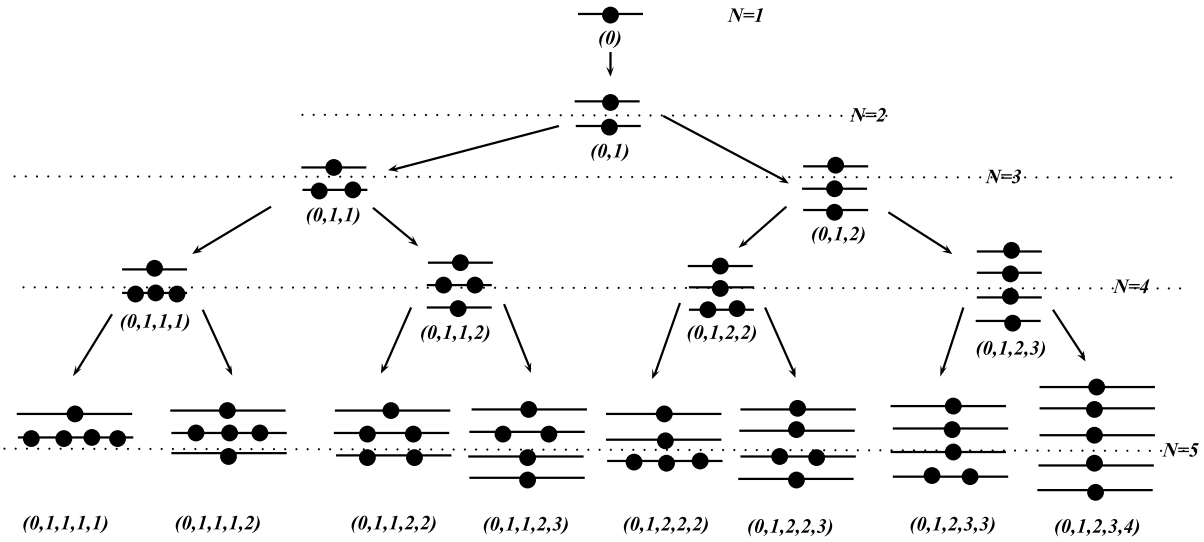


Figura 4.6: Cada nivel corresponde a las posibles distribuciones para un N dado, para esta figura el ultimo nivel corresponde a $N = 5$ nodos.

un solo nodo en el $Nivel$ 0:

$$\frac{2^{N-1}}{2} = 2^{N-2} \tag{4.13}$$

Otro punto importante a destacar del árbol binario de la Figura 4.5 en su parte correspondiente a redes jerárquicas, es que al construirlo no solo generamos las posibles distribuciones para redes jerárquicas de N nodos; las cuales corresponden a los caminos del nodo raíz hasta las nodos en el $Nivel$ $N-1$, sino que también están presentes todas las posibles distribuciones para redes jerárquicas formadas por un número menor de nodos. Si los caminos hacia los nodos del $Nivel$ $N-1$ corresponde a las distribuciones de N nodos, el nivel $N-2$ corresponde a distribuciones de $N-1$ nodos, hasta llegar al primer $Nivel$ que sería el caso de un solo nodo, como se puede ver en la Figura 4.6.

Ahora que sabemos que forma tienen las distribuciones en niveles de los nodos en redes jerárquicas y cuantas son, solo nos faltaría conocer como se comporta el índice de desigualdad, Id en estas configuraciones, esto lo haremos mas adelante, pero antes definiremos otro índice.

4.5. Índice de desigualdad externo.

Como hemos visto algunas configuraciones de los nodos presentan valores similares, para su indice de desigualad Id para distinguir dos redes jerárquicas de una manera mas clara, introducimos una cantidad basada en el **centro de masa**, que llamaremos

índice de desigualdad externa (Ie). Mientras que el índice de desigualdad Id la podemos interpretar como una medida de que tan cercanos se encuentran los nodos mutuamente, el índice de desigualdad externo Ie nos ayudara a a distinguir en torno a que nivel se encuentran la mayor concentración de los nodos. Como vimos en secciones previas, la distancia para los nodos en redes jerárquicas, quedo definida como el número de vínculos necesarios para llegar del nodo raíz al nodo deseado. Tener una definición de distancia nos permite extender varios conceptos y la de centro de masa (Ie) para redes jerárquicas no es la excepción, este quedaría como:

$$Ie = \frac{1}{N} \sum_{i=1}^{l_{max}} Oc_i l_i, \tag{4.14}$$

donde N es el número de nodos total, l_{max} es el nivel máximo de la red jerárquica, Oc_i es la ocupación del $i'simo\ Nivel$, es decir, el número de nodos en el $Nivel\ i$ y l_i es el $Nivel\ i$.

4.6. Índice de desigualdad para algunas redes jerárquicas

Para esta sección vamos a utilizar dos procedimientos para analizar el comportamiento del Id en distintas distribuciones en niveles. El primero de estos consiste en determinar el Id para todas las posibles distribuciones en niveles de un número N de nodos dado, para esto nos apoyaremos en la representación de las distribuciones como $N'adas$ que vimos antes. Para el segundo procedimiento, elegimos algunas redes jerárquicas de topologías conocidas, como son, los arboles regulares, arboles aleatorios y arboles generados a partir del principio de vinculación preferencial. Para ambos casos incrementamos el número de nodos de estas estructuras y calculamos el valor de Id para cada uno de ellos. Para dichos procedimientos construimos los programas necesarios para determinar el valor del índice de desigualdad Id.

4.6.1. Índice de desigualdad para distintas distribuciones en niveles

Retomando la idea de $N'ada$ como representación de las distintas distribuciones posibles para redes jerárquicas, determinemos el valor de Id (ecuación 4.10) de cada una de ellas para algunos valores de N. Comencemos con el ejemplo que hemos utilizado en secciones anteriores de $N = 5$, para el cual queda de la siguiente forma, ver Figura 4.7. Como podemos observar el Id se incrementa a medida que el número de niveles crece y la distribución de nodos en los distintos niveles se vuelve cada vez más homogénea; hasta llegar al caso extremo en que solo hay un nodo por nivel, donde $Id = 1$.

Ahora, puesto que el número de posibles distribuciones crece de manera exponencial (ecuación 4.13 sería poco practico calcular los valores de Id de las distribuciones de un número mayor de nodos de forma directa. Para esta tarea elaboramos un programa (ver Apéndice A), el cual pide al usuario el número de nodos a distribuir en niveles y

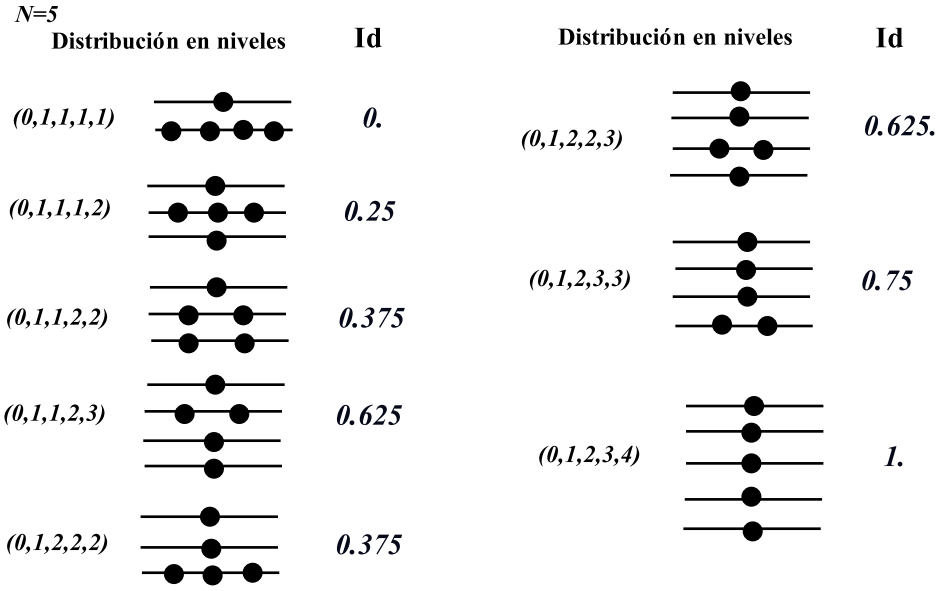


Figura 4.7: En esta figura se pueden observar todas las posibles distribuciones en $Niveles$ para $N = 5$ nodos. Se muestran dos grupos de columnas, en cada una de ellas, a la izquierda esta la $N'ada$ asociada a la distribución de la derecha y el Id que se obtuvo para cada distribución.

regrese un conjunto de $N'adas$ correspondientes a todas las posibles distribuciones, ,)En el Apéndice A

4.6.2. Índice de desigualdad para estructuras dadas de redes jerárquicas

Id para arboles regulares

Existen distintas aplicaciones donde podemos encontrar arboles regulares, como por ejemplo , problemas de búsqueda, ordenamientos o codificación. Llamamos arboles regulares a los arboles con raíz que tienen la propiedad de que todos sus nodos internos tienen el mismo número de hijos. Para ser mas claros, a continuación damos la definición que se usa en teoría de gráficas para árbol $m - ario$ término que junto a árbol regular usaremos indistintamente.

Definición 4.6.1. *Llamamos árbol $m - ario$ a un árbol con raíz, si cada nodo interno tiene no más de m nodos hijos. Se le llama árbol $m-ario$ completo si cada nodo interno tiene exactamente m nodos hijos.*

Un caso particular a menudo utilizado es el árbol con $m = 2$ el cual es llamado árbol binario. El la figura 4.11 se pueden ver algunos ejemplos de arboles $m - arios$.

Para determinar el comportamiento del Id para redes jerárquicas que tienen una topología de árbol regular, generamos arboles $m - arios$ con m que varia de 2 hasta 10,

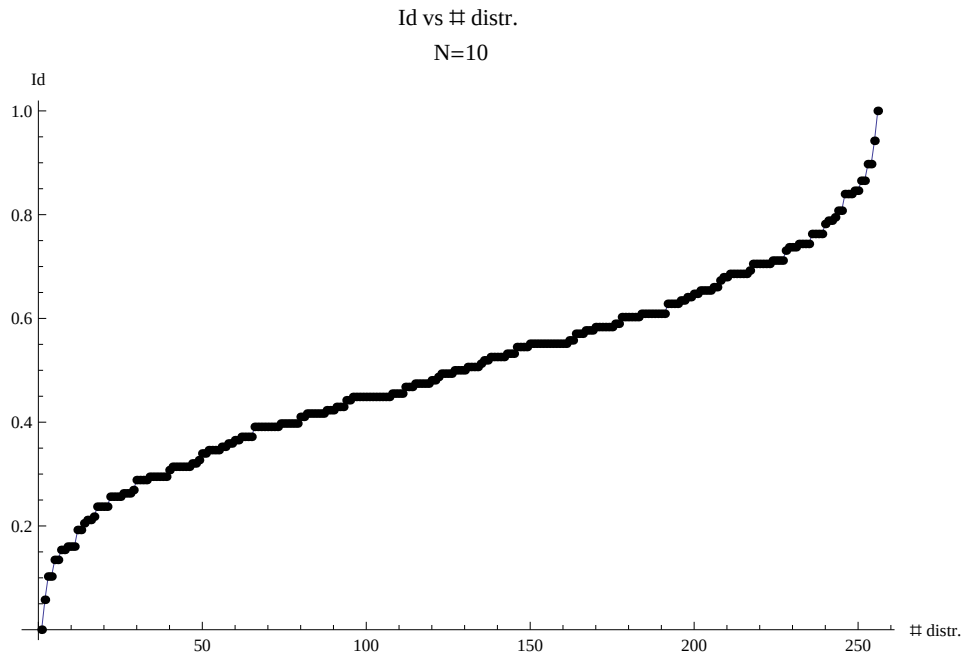


Figura 4.8: Id para todas las $N'adas$ de N=10 nodos, ordenadas en forma creciente

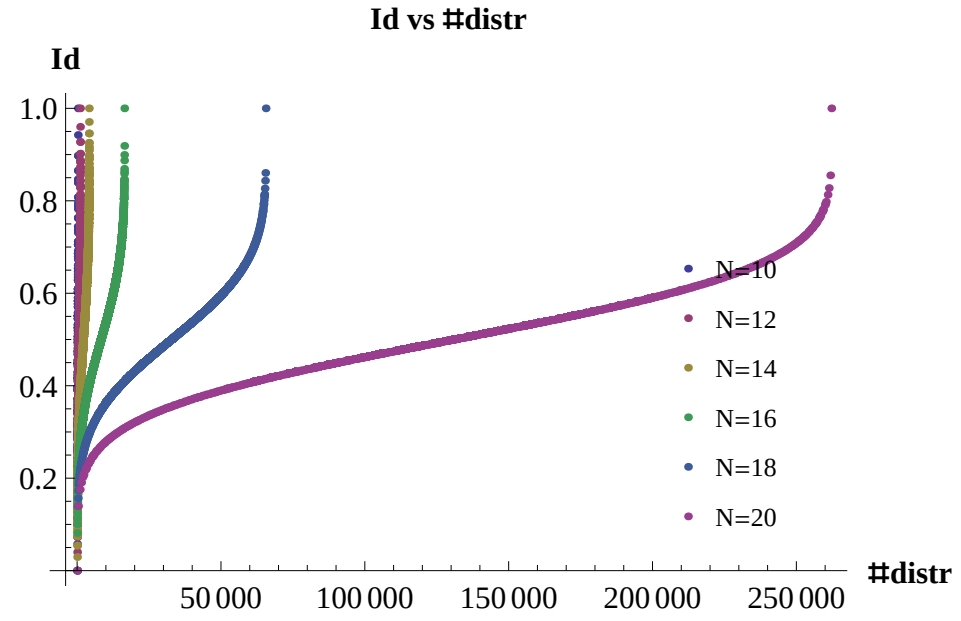


Figura 4.9: Id para todas las $N'adas$ de N=10,12,14,16,18,20 nodos, ordenadas en forma creciente.

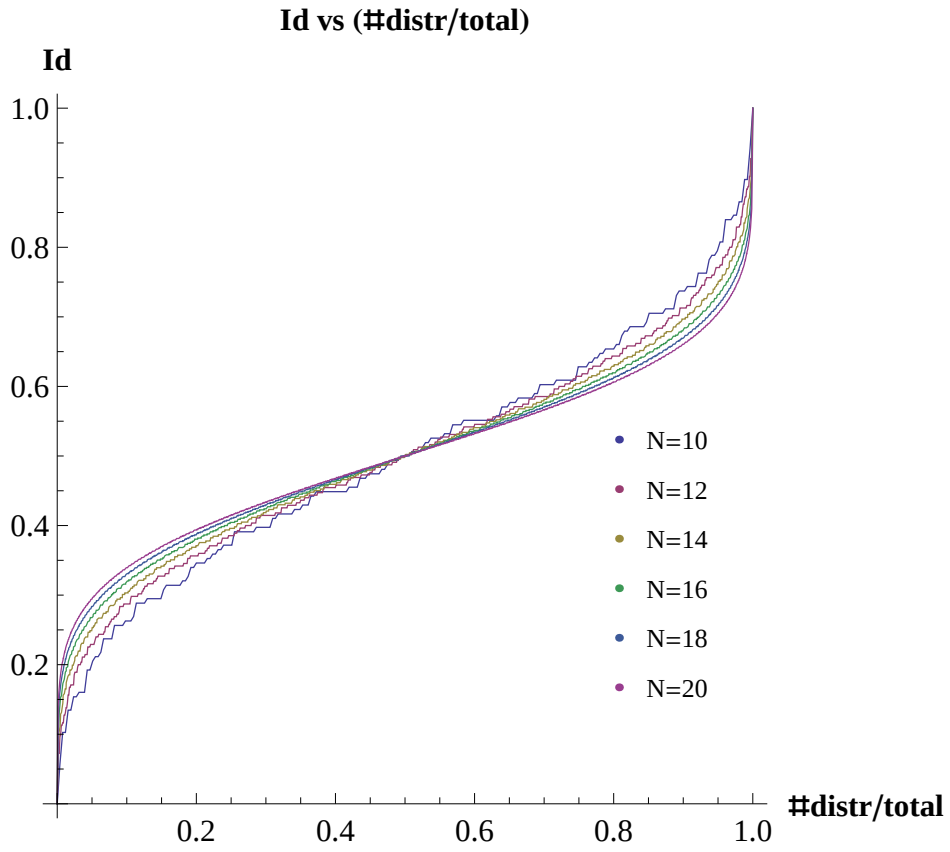
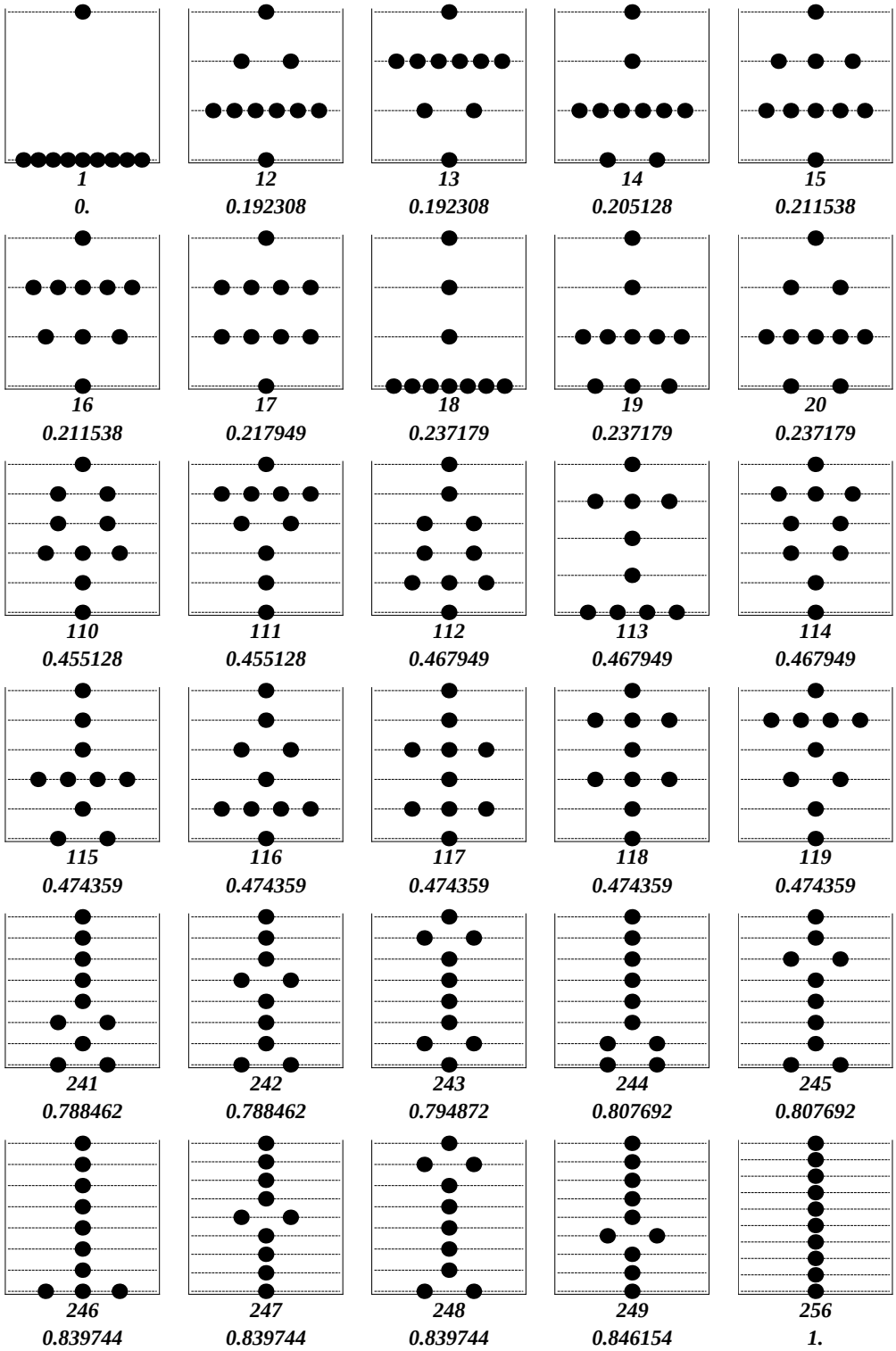


Figura 4.10: Id para todas las $N'adas$ de N=10,12,14,16,18,20 nodos, normalizada por el numero de distribuciones.



Cuadro 4.1: Algunas distribuciones de $N = 10$ nodos, debajo de cada figura se muestran dos números el primero es la posición de la distribución en la gráfica ordenada de forma ascendente y el segundo es el Id.

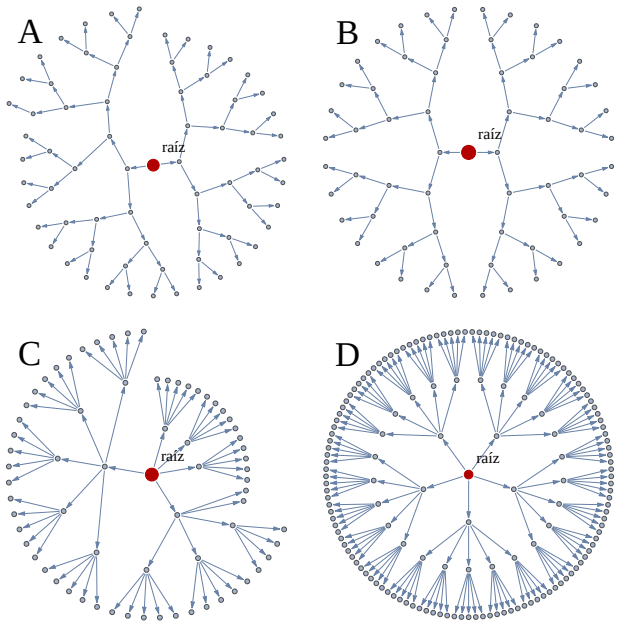


Figura 4.11: Se presentan cuatro arboles $m - arios$. A es un árbol binario y B es un árbol binario completo. C corresponde al caso m=5 y D tiene el mismo valor (m=5) pero completo.

incrementando el número de nodos (N) de 3 a 3000 para cada uno de ellos, en pasos de 10. Para todos estos determinamos el valor de Id obteniendo la figura 4.12. Como podemos observar los valores que se obtienen para dichas estructuras cuando (N) es pequeño llegan a alcanzar valores de Id que se encuentran entre $[0.1, 0.2]$, sin embargo, conforme se incrementa el número de nodos N la gráfica se vuelve asintótica a 0. Para una mejor visualización del comportamiento se obtiene la gráfica en escala logarítmica. Ver figura 4.13. Se puede observar claramente como el árbol binario tiene valores de Id por encima del resto de los demás arboles regulares.

Id para arboles aleatorios

Los arboles aleatorios han sido utilizados en distintos contextos, como por ejemplo: modelar la propagación de epidemias, investigar y construir arboles familiares, en el comportamiento de *cartas cadena* o en juegos tipo pirámide. Estos arboles los podemos considerar como el resultado de un proceso evolutivo, veamos como es esto. Supongamos que comenzamos con un solo nodo, dicho nodo lo designamos como el nodo raíz de la estructura jerárquica resultante y este primer nodo es etiquetado con el número 0, en el primer paso vincularemos un nodo etiquetado con el número 1 con el único nodo presente en la estructura en ese paso (nodo 0), en el siguiente paso vincularemos el nodo con etiqueta 2 con los nodos presentes en este paso, hay dos posibilidades, las cuales podrían ser, el nodo 0 (raíz) o el nodo 1, en el siguiente paso vincularíamos al

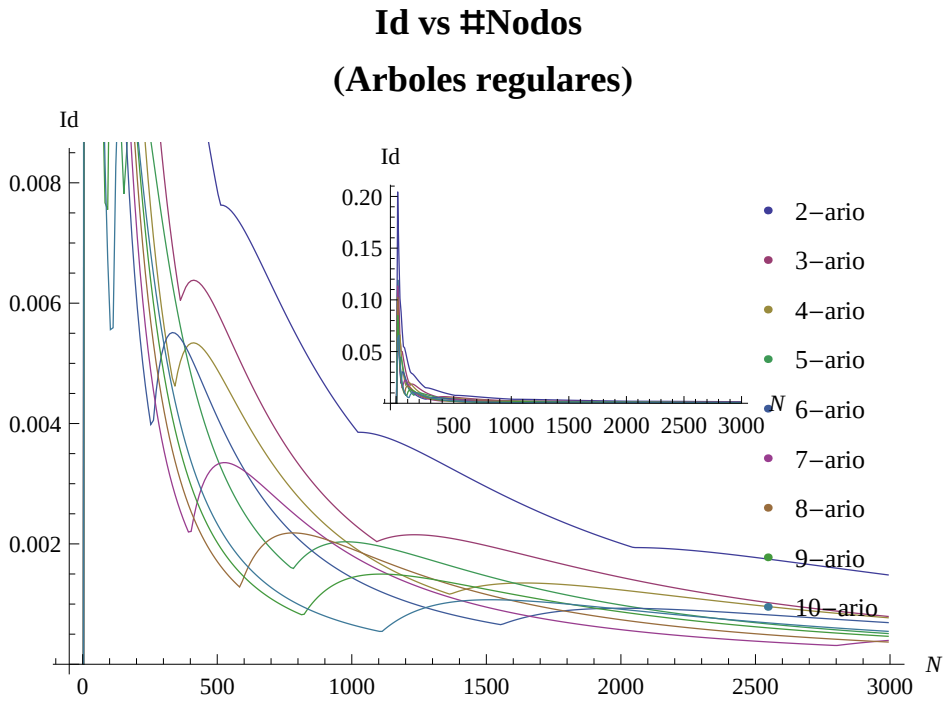


Figura 4.12: Se gráfica el valor de Id para distintos arboles regulares, para los cuales se incrementa el número de nodos N de 3 hasta 3000 en pasos de 10. El gráfico interno muestra todo el rango de valores que toma Id. Los valores de m van de 2 hasta 10.

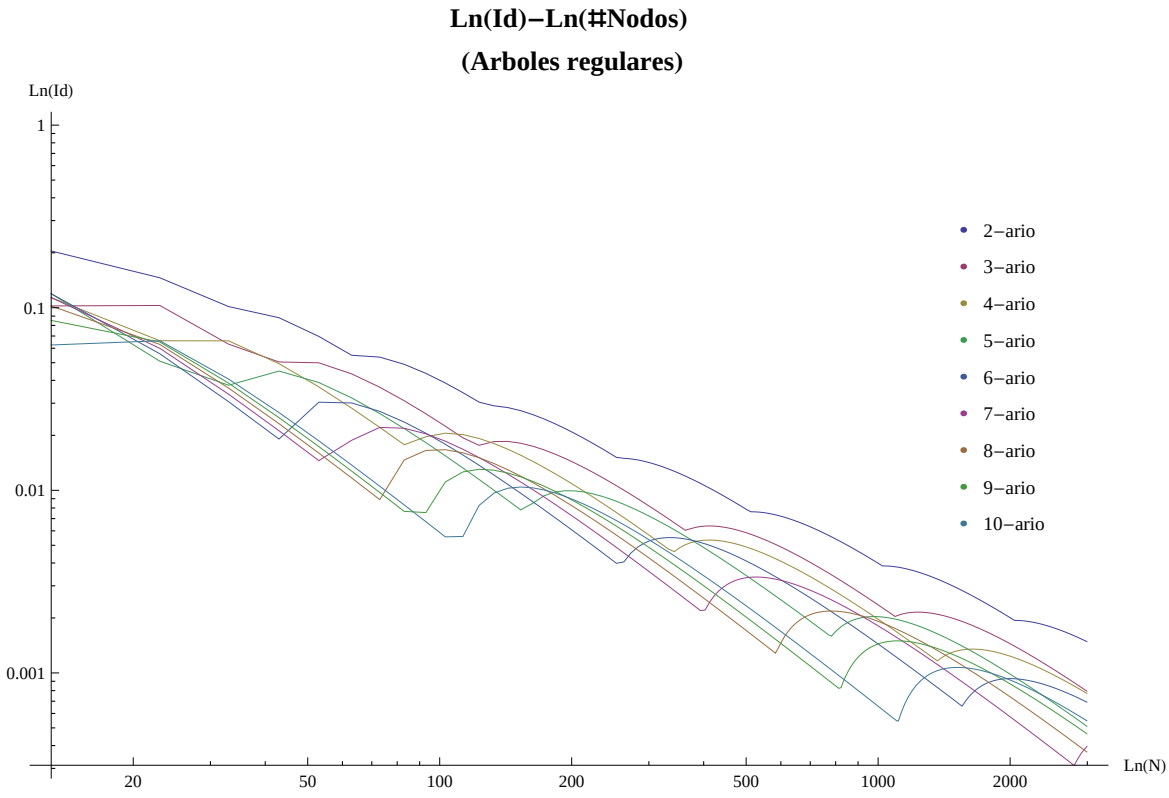


Figura 4.13: Gráfica en escala logarítmica de redes jerárquicas regulares.

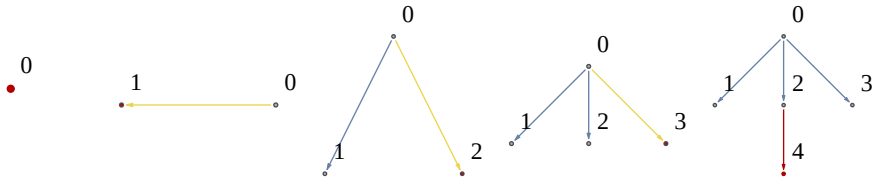


Figura 4.14: De izquierda a derecha el arbol aleatorio al $t = 0$, $t = 1$, $t = 2$, $t = 3$ y $t = 4$. El nodo y el vínculo con un color distinto al azul corresponden a los que se agregan en cada paso.

nodo con etiqueta 3 con alguno de los tres nodos existentes en ese paso (nodos 0, 1 o 2). Y de esta forma sucesivamente. En general, cuando tenemos un árbol con las etiquetas $1, 2, 3, \cdots, i$ vinculamos el nodo con la etiqueta $i+1$ a alguno de los nodos ya existentes. Entonces, como existen exactamente i formas de vincular el nuevo nodo con la etiqueta $i + 1$, decimos que existen exactamente $(N - 1)!$ de posibles arboles de N nodos. La forma mas natural en que podemos pensar la distribución de probabilidad de los arboles aleatorios de N nodos, es asumir, que cada uno de los $(N - 1)!$ es igualmente posible. Esta distribución también puede obtenerse a partir de una dinámica que realice este proceso evolutivo, es decir, que vincule sucesivamente un nodo a cualquiera de los nodos ya existentes, todos con la misma probabilidad de ser el padre de este nuevo nodo, (ver figura 4.14), hasta llegar al número N de nodos deseados. Con estas sencillas reglas podemos construir un programa que genere arboles jerárquicos aleatorios para posteriormente analizar las características deseadas.

Entonces, para obtener el comportamiento de Id para las redes jerárquicas aleatorias construimos el programa (Apéndice B) antes descrito, nuevamente para obtener una estadística útil para representar el comportamiento de Id generamos 1000 redes jerárquicas aleatorias para un numero N de nodos dado. Comenzamos con un número de nodos $N = 3$ hasta llegar a $N = 2000$ a pasos de 10 nodos. Para nuestro objetivo las etiquetas de los nodos solo tienen importancia para la construcción del programa, sin embargo, es importante señalar que la numeración que se obtiene podría ser importante si se desea identificar la edad de los nodos en este tipo de estructuras, valores mas grandes en las etiquetas indicarían nodos más jóvenes mientras que valores más pequeños indican nodos mas viejos. Antes de mostrar la gráfica obtenida para las redes jerárquicas aleatorias, construiremos las redes jerárquicas con el principio de vinculación preferencial que es el tema de la siguiente sección y compararemos los dos resultados.

Id para arboles con vinculación preferencial

En las últimos años el estudio de redes que presentan una vinculación distinta de la equiprobable ha tomado gran fuerza, sobre todo porque muchas de las redes reales muestran características que no se explican con este principio. La vinculación preferencia ha aparecido repetidamente en Matemáticas como en las Ciencias Sociales con distintos nombres, solo por citar algunos ejemplos tenemos: George Udmy Yule (1871-

1951) en 1925 uso el principio de vinculación preferecial par explicar la distribución de ley de potencias del número de especies por genero de algunas plantas, este proceso lleva por nombre proceso de Yule. Rober Gibrat (1904-1980) en 1931 propone que el tamaño y la tasa de crecimiento de una compañía son independientes. Puesto que grandes firmas crecen más rápido. Se le llama crecimiento proporcional, la cual es una forma de vinculación preferencial. George Kinsley Zipf (1902-1950) en 1941 utilizo el principio de vinculación preferencial para explicar la distribución de cola larga de la riqueza en la sociedad. En 1999 se introduce el término vinculación preferencial (preferential attachment) en el articulo de Barabási y Albert para explicar la ley de potencias en las redes.

Dos son las características que señala Barabási que diferencian las redes reales de las aletorias son: **Crecimiento**. Mientras que en los primero modelos de redes aleatorias en número de nodos N era fijo las redes reales son el resultado de un proceso de crecimiento que continuamente incrementa N. **Vinculación preferencial**. Mientras los nodos en las redes aleatorias eligen de manera equiprobable un nodo para vincularse, en las redes reales los nodos prefieren vincularse a los nodos con mas vínculos. Estas dos propiedades coexisten en la redes reales y llevan a introducir un modelo mínimo capaz de generar redes con una distribución de ley de potencias en el grado de sus nodos.

Para construirlo este modelo se comienza con m_0 nodos los cuales se eligen arbitrariamente, siempre y cuando cada nodo tiene al menos un enlace. Los pasos para construir una red son los siguientes:

Crecimiento.

En cada paso se agrega un nuevo nodo con m ($\leq m_0$) vínculos que conectan al nuevo nodos a m nodos que están presentes en la red.

Vincualcion preferencial.

La probabilidad $\prod(k)$ que uno de los vínculos del nuevo nodo conecte a el nodo i dependen del grado k_i del nodo i como:

$$\prod(k) = \frac{k_i}{\sum_j k_j} \tag{4.15}$$

Este modelo recibe el nombre de modelo de Barabási-Albert.

Para generar redes jerárquicas con el principio de vinculación preferencial es suficiente con que $m_0 = 2$ nodos, con un vínculo entre ellos, uno de estos nodos se etiqueta con el valor 1 y el otro con el valor 2 (como en la sección anterior la etiqueta solo es importante para el programa). Los nodos que se van agregando a esta red tienen un valor de $m = 1$ que eligen vincularse con los nodos que se encuentran en la estructura con probabilidad dada por la ecuación 4.15. Nuevamente generamos a partir de un programa (Apéndice C) 1000 redes con un número de nodos N dado, el cual incrementamos en pasos de 10 a partir de 3 hasta 2000 nodos. Para cada uno de estas estructuras calculamos el valor de Id y obtenemos como incertidumbre para cada valor su desviación estandard. La siguientes figuras 4.15 y 4.16 muestra los resultados obtenidos.

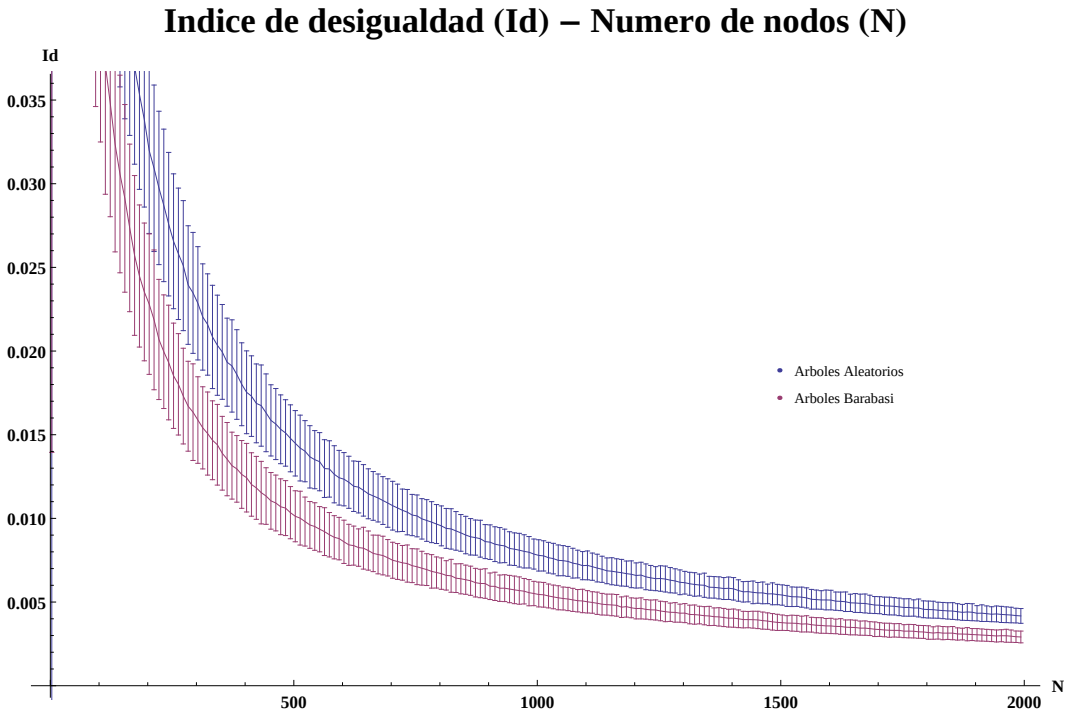


Figura 4.15:

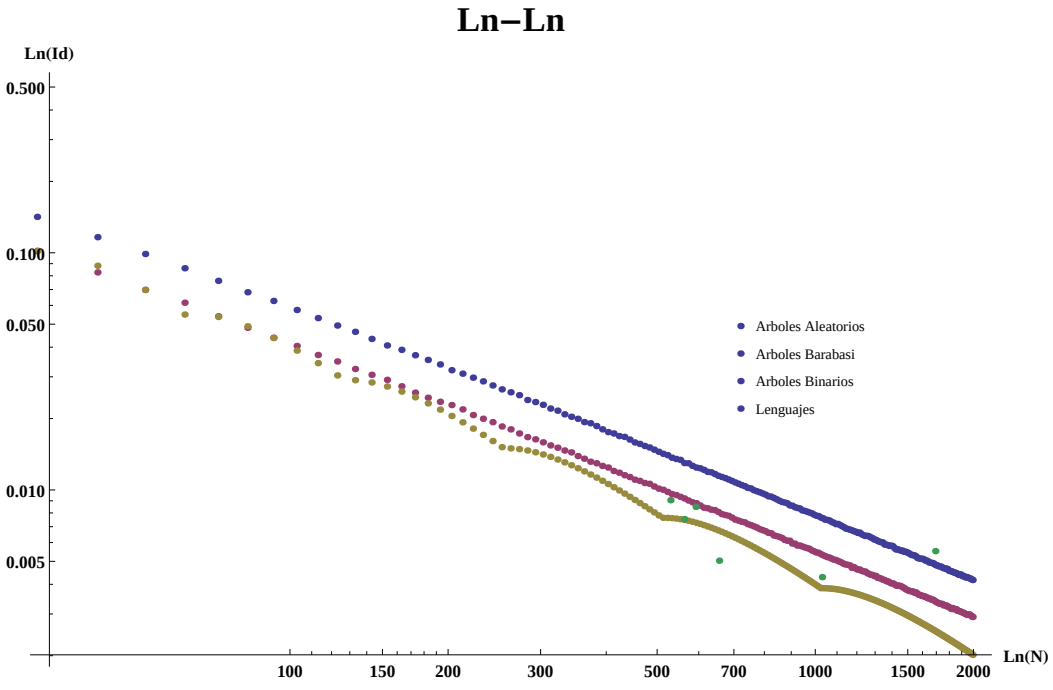


Figura 4.16:

Capítulo 5

Clasificación jerárquica de los lenguajes

"Si el Homo Sapiens ha logrado permanecer, colonizar y expandir su presencia en la Tierra, ello se debe a su habilidad para reconocer y aprovechar los elementos y procesos del mundo natural, un universo caracterizado por una característica esencial: la diversidad."

Víctor M. Toledo y Narciso Barrera-Bassols

("Memoria Bioculatural")

En este capítulo obtenemos algunas características cuantitativas de los árboles de clasificación taxonómica de las cinco familias lingüísticas mas grandes (considerando el número de lenguajes pertenecientes a cada una de ellas) del mundo; las cuales abarcan mas del 85 % del total de hablantes en el mundo. Determinamos los distintos índices propuestos en capítulo 4 y obtenemos la distribución de grado para cada una de las familias lingüísticas.

5.1. Introducción

La población mundial es poco mas de 7,000 millones[1] al día de hoy, sin embargo, dicha población puede ser agrupada en cerca de 7,105 comunidades distintas respecto a su lenguaje materno[2], un número modesto si se considera que llego a ser de 12,000 antes de la expansión colonial europea en el siglo xv [memoria bioculatural pag18]. Estos lenguajes pueden ser clasificados de distintas formas dependiendo el conjunto de criterios utilizados, lo cual nos permite analizar relaciones útiles entre ellos, algunos ejemplos de estas clasificaciones son:

. **Por el lugar en donde se hablan**, de esta forma podemos referirnos a las lenguas que se hablan en Chiapas, Oaxaca o en México incluso a una escala mas grande,

[1] 7,156,695,000 según el U.S. Census Bureau el dia lun 31 mar 2014 18:05:59 CST http://www.census.gov/popclock/

[2] https://www.ethnologue.com/

los lenguajes que se hablan en África, Asia, América, etc. Esta clasificación también permite distinguir por grupos étnicos que en algunos casos son mas grandes que las fronteras de estados, países o continentes.

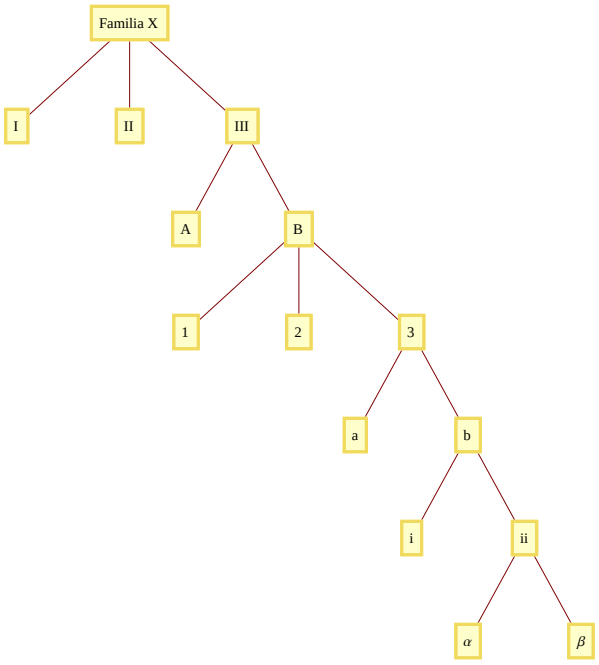
. **Clasificación por las características del lenguaje**, algunas de estas pueden ser por el orden de las palabras, sujeto, objeto y verbo. Por la distinción de significado por la entonación, o de acuerdo al sistema de sonidos, el numero de vocales y consonantes que se utilizan.

. **Clasificación genética**. Esta clasificación intenta agrupa respecto al lenguaje ancestral común.

Esta última la más importante para los fines de este trabajo, nos permite analizar el proceso de diversificación que ha dado lugar a las lenguas que conocemos actualmente y la cual ha sido reforzada por su correspondencia con las genealogías genéticas (Cavalli-Sforaza, 2001), lo cual nos permite destacar que de la misma forma en que las poblaciones humanas se diversificaron y evolucionaron, también los lenguajes lo hicieron. Para ejemplificar este proceso, podemos mencionar las lenguas romances (Español, Francés, Italiano, Portugués, Rumano, etc) grosso modo y sin entrar en los detalles y las controversias entre lingüistas, podemos decir, que durante el período en que se extendió el imperio romano por gran parte de Europa y mantuvo una red de caminos que hacia sencillo poderse mover por el imperio, soldados, mercaderes y gente común se desplazo por todos estos dominios llevando consigo no solo sus costumbres sino también su lengua: el Latín. Conforme paso el tiempo la gente no hablante de Latín lo fue adoptando y el Latín hablado en todo el imperio fue probablemente bastante similar. A la caída del imperio Romano la movilidad por la red de caminos se volvió mas difícil y peligrosa, el habla de distintas regiones como Francia, Italia o España se comenzó a separar, primero como pequeñas diferencias en el acento y el vocabulario (como las que caracterizan al español de España y el español de México), con el transcurso del tiempo estas se convirtieron en diferencias de vocabulario, pronunciación e incluso de sintaxis, sin descartar, la influencia de pueblos vecinos de habla no Latina. A pesar de todo esto, idiomas como el Francés el Italiano o el Español entre otras, actualmente mantienen bastantes similitudes, lo que se debe a que se desarrollaron a partir de una misma "proto-lengua" que era el Latín. Entonces decimos que el lenguaje que desciende de la lengua de los romanos, están relacionadas genéticamente y las llamamos, lenguas romances (o románicas).

Los lingüistas han demostrado que casi todos los idiomas pertenecen a alguna familia de lenguaje, cada una de las cuales desciende de algún protolenguaje ahora perdido Entre estas familias se encuentran Sino-Tibetan, Niger-Congo, Austronesian y Afro-Asiatica. sarSin embargo esta clasificación no esta excenta de controversia

Distribución del numero de hijos para la clasificacion taxonomica de los lenguajes Desarrollo A partir de la información que proporciona la pagina de Ethologue se construyen los arboles genealogicos de las 5 familias mas grandes de lenguajes, las cuales abarcan cerca del 85Esto se puede deber a las caracteristicas geograficas (islas), aislamienteo lo cual permite una gran variedad en el numero de lenguas.



Familia x
I Rama Principal
II Rama Principal
III Rama Principal
A Grupo
B Grupo
1 Subgrupo
2 Subgrupo
3 Subgrupo
a Rama
b Rama
i Subrama
ii Subrama
α Grupo
β Grupo

Cuadro 5.1: A la izquierda se representa una estructura de árbol genético para una familia hipotética. A la derecha la misma estructura de árbol en su representación por un sistema de sangría.

5.2. Procedimiento

La clasificación genética de los lenguajes en muchas ocasiones es representada en forma de árbol con raíz, donde el primer nodo de esta jerarquía es considerada la Familia. Otra forma de representar estas estructuras que aparecen en los libros y paginas de internet especializadas en clasificación de los lenguajes, es por medio de una sistema de sangrado, donde cada nivel de ramificación sucesivo se indica por una sangría a la derecha, ademas de que cada nodo es precedido por una letra o número de identificación del nivel taxonómico o grupo. De esta forma una diagrama de árbol puede ser representado en forma de listado, como se puede ver en el cuadro 5.1.

Para la construcción de los árboles de clasificación de las seis principales familias lingüísticas que utilizamos en este trabajo, nos basamos en los datos obtenidos en la pagina http://www.ethnologue.com/, la cual es considerada la mas grande base de datos en el tema. La forma en que se presentan los datos en Ethnologue es por medio del sistema de sangría, mostramos una parte de la familia Indo-Europea en la figura 5.1. En el nivel más alto se encuentra el nombre de la familia lingüística y a la derecha

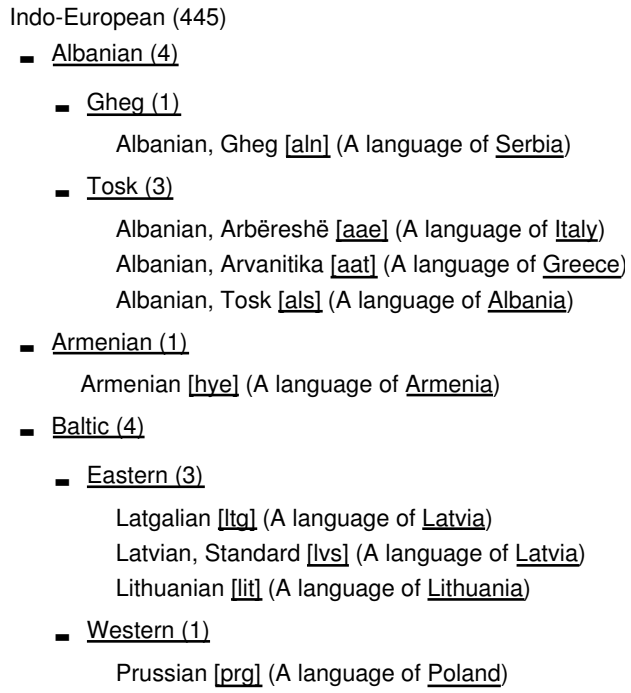


Figura 5.1: Fragmento de información para la familia Indo-Europea. En esta figura podemos ver las primeras (en orden alfabético) tres ramas principales de la familia Indo Europea (Albanian, Armenian, Baltic). El último nivel en esta representación, la cual corresponden a los renglones más a la derecha, son los nombres de las lenguas actuales, a su derecha entre corchetes se encuentra la clave con que dicho lenguaje es identificado en esta base de datos. Tomado de: http://www.ethnologue.com/subgroups/indo-european.

de está, entre paréntesis el número de lenguajes actuales, en el siguiente nivel cada subgrupo nuevamente esta acompañado de un paréntesis con el número de lenguajes que pertenecen al subgrupo, y de esta forma hasta llegar al último nivel, el cual corresponde a los nombres de los lenguajes actuales, estos están acompañados a su derecha entre corchetes por una clave que identifica a cada lenguaje con su genealogía, por ejemplo: si nosotros buscamos en Ethnologue [hye] nos dirige a una página con la información del Armenio (Armenian) en la cual encontraremos su clasificación como Indo European - Armenian.

Aunque estos datos son sumamente valiosos por si mismos y sin los cuales no podríamos trabajar, es necesario representarlos de una forma apropiada para poder extraer información acerca de las propiedades de estas estructuras, es decir, es necesario construir las redes jerárquicas con raíz de estas familias lingüísticas. Por lo tanto, a partir de los listados arriba mencionados, nos dimos a la tarea de construir los árboles genealógicos de cada una de las seis familias lingüísticas mas grandes (desde el punto de vista del número de lenguajes pertenecientes a cada una ellas), estas familias lingüísticas son: Familia Afro-Asiatica, Familia Austronesia, Familia Indo-Europea, Familia Niger-

Congo, Familia Sino-Tibetana y Familia Trans-Neoguineana. Las redes jerárquicas que obtuvimos para cada una de estas familias se presentan en el cuadro 5.2.

Ahora toda la información contenida en los listados tiene una representación de red jerárquica con raíz, a partir de la cual, podemos calcular los índices y medidas que propusimos en el capitulo 4, como también, aplicar algunas de las herramientas existentes para el análisis de redes, como la distribución de grado de los nodos. Este es el objetivo de las siguientes secciones.

5.3. Resultados

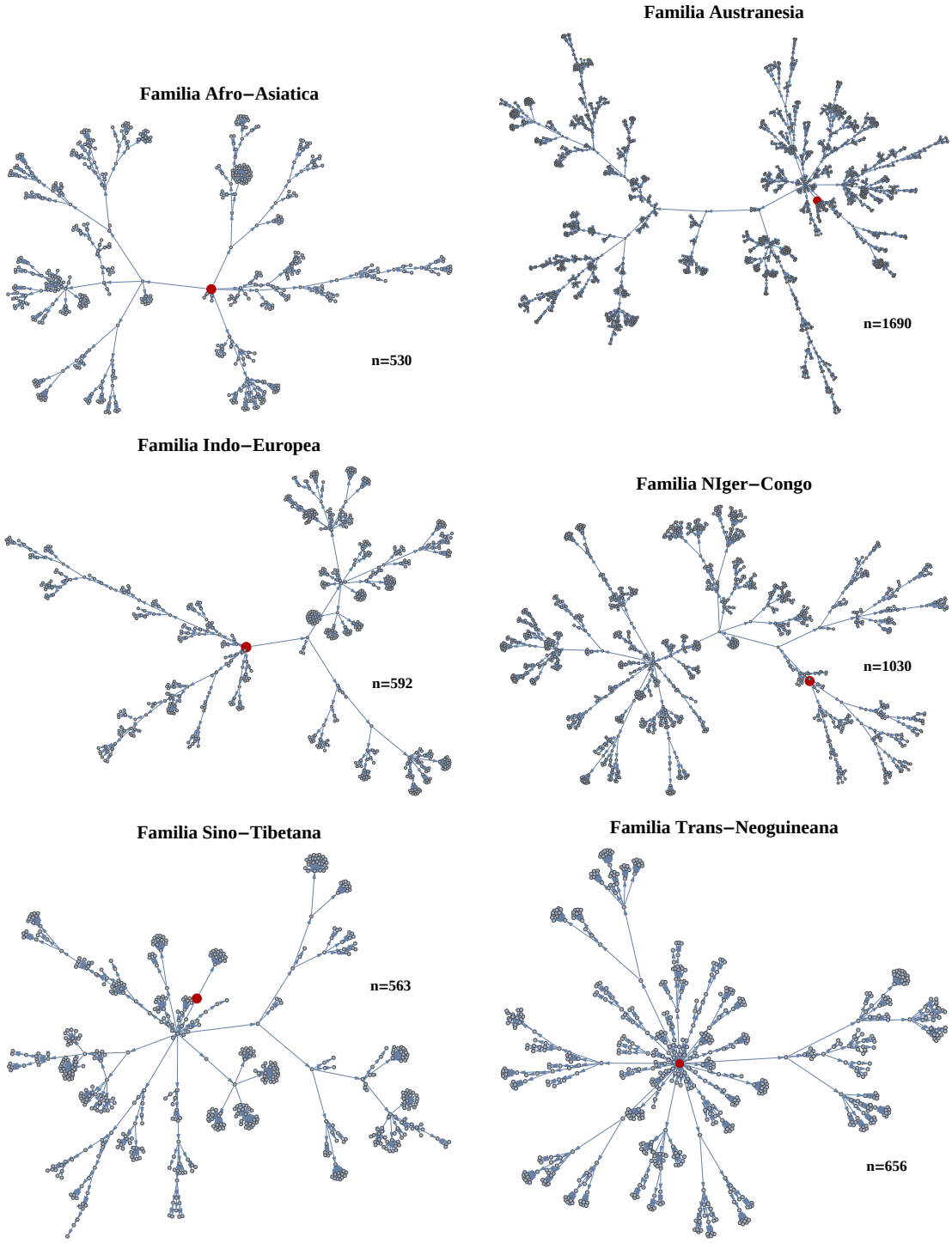
5.3.1. índices de desigualdad de los lenguajes

A partir de estas redes podemos asociar a cada nodo un nivel, que es la distancia que hay desde el nodo padre a nodo deseado, en el cuadro de las redes se puede observar en rojo el nodo origen o nodo padre. Con estos datos determinamos los valores de de los indices Id e Ie, índice de desigualdad e índice de desigualdad externo respectivamente, que definimos en el capítulo 4 y cuyas expresiones son:

$$Id = \frac{3}{(N+3)(N-2)(N-1)} \left[\sum_{i \neq j=1}^{N} |n_i - n_j| - 2(N-1) \right] \qquad \forall N > 2.$$

$$Ie = \frac{1}{N} \sum_{i=1}^{l_{max}} Oc_i l_i,$$

Además de determinar los índices de desigualdad determinamos los valores de la distancia media $\langle d \rangle$, como también, los de la distancia máxima d_{max} en cada red, los valores obtenidos para cada una de las redes lingüísticas se comparan con los valores obtenidos para redes jerárquicas aleatorias y redes jerárquicas creadas con el principio de vinculación preferencial del mismo número de nodos, obteniendo el cuadro 5.3.



Cuadro 5.2: Se muestran las seis principales familias lingüísticas, estas familias representan poco mas del 85 % de los hablantes del mundo y poco mas de 63 % de las lenguas vivas actualmente. En rojo podemos observar el nodo raíz y n representa el número de nodos que conforma cada red.

Familia	**n**	$\langle d \rangle$	d_{max}	I_d	I_e
Afro-Asiatica	530	4.954717 ± 1.451753	8	0.009053	4.954717
Árbol Aleatorio	530	5.852977 ± 0.593371	12.405333 ± 1.509900	0.0138784 ± 0.001867	5.852977 ± 0.001867
Árbol Barabási	530	3.918857 ± 0.570609	8.873333 ± 1.231988	0.009706 ± 0.001470	3.918857 ± 0.001470
Austronesia	1690	7.427811 ± 2.737985	14	0.005521	7.427811
Árbol Aleatorio	1692	6.999175 ± 0.586811	15.182333 ± 1.553456	0.004835 ± 0.000530	6.999175 ± 0.000530
Árbol Barabási	1692	4.492224 ± 0.581097	10.661 ± 1.321	0.003379 ± 0.000421	4.492224 ± 0.000421
Indo-Europea	592	5.104730 ± 1.574708	10	0.008484	5.104730
Árbol Aleatorio	592	5.941572 ± 0.577684	12.599 ± 1.516	0.012462 ± 0.001624	5.941572 ± 0.001624
Árbol Barabási	592	3.973484 ± 0.572422	9.006667 ± 1.212619	0.008745 ± 0.001262	3.973484 ± 0.001262
Niger-Congo	1030	6.286408 ± 1.376080	8	0.004289	6.286408
Árbol Aleatorio	1030	6.506528 ± 0.600607	13.964333 ± 1.541164	0.007610 ± 0.000923	6.506528 ± 0.000923
Árbol Barabási	1030	4.262970 ± 0.572939	9.915333 ± 1.263605	0.005352 ± 0.000730	4.262970 ± 0.000730
Sino-Tibetana	563	4.975133 ± 1.322978	8	0.007534	4.975133
Árbol Aleatorio	563	5.903469 ± 0.571715	12.56 ± 1.48	0.013111 ± 0.001712	5.903469 ± 0.001712
Árbol Barabási	563	3.959650 ± 0.575474	8.936333 ± 1.243434	0.009178 ± 0.001365	3.959650 ± 0.001365
Trans-Neoguineana	656	3.254573 ± 1.022261	5	0.005031	3.254573
Árbol Aleatorio	656	6.039608 ± 0.580395	12.877333 ± 1.513181	0.011406 ± 0.001464	6.039608 ± 0.001464
Árbol Barabási	656	4.037389 ± 0.566890	9.222 ± 1.253	0.008036 ± 0.001169	4.037389 ± 0.001169

Cuadro 5.3: Se

5.3.2. Hojas y nodos

En una red tipo árbol con raíz, podemos clasificar los nodos en dos tipos considerando su grado de salida. Si el grado de salida del nodo es cero, entonces lo llamamos hoja y si es distinto de cero simplemente nodo o nodo interno. Las hojas son de bastante importancia en una estructura genealógica, puesto que su número representan los descendientes finales generados por la evolución de la estructura.

Si consideráramos que el objetivo principal de un árbol genealógico, es obtener un número considerable de hojas, con un número mínimo de nodos, se nos plantearía el problema de como medirlo.

5.3.3. Cociente hojas y nodos

La forma en que proponemos medir que tan efectiva (coeficiente de efectividad, e) es una estructura tipo árbol con raíz para generar hojas es por el siguiente cociente:

$$e = \frac{h}{n} \tag{5.1}$$

donde h es el número de hojas total en un árbol y n el número de nodos total que lo conforma. Por ejemplo en el caso de un árbol aleatorio, el cociente es $e = 0.5$, es decir, la mitad de los nodos son nodos internos del árbol y la otra mitad son hojas. En un árbol construido con vinculación preferencial obtenemos que el cociente es $e = 0.6$, es decir, 40 % de los nodos son nodos internos de la estructura y 60 % son hojas. Nuestra siguiente tarea por lo tanto sería determinar el valor de e para nuestras árboles genealógicos.

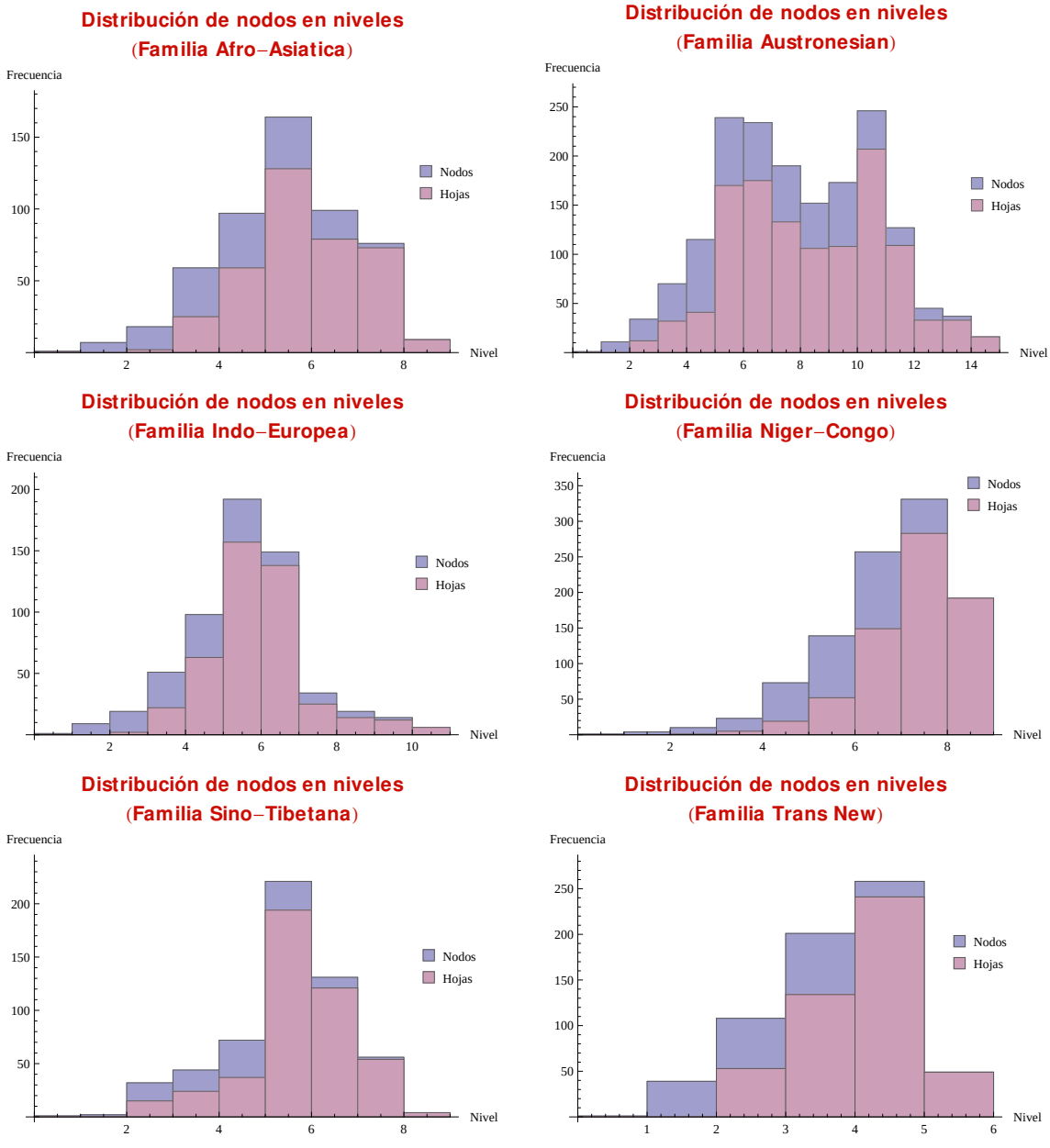
e en la clasificación genética de los lenguajes.

A partir de los árboles construidos para las seis principales familias lingüísticas y el programa elaborado para determina e (el cual se anexa en el APENDICE), determinamos el cociente para cada árbol obteniendo el cuadro 5.4.

Como podemos observar en la cuadro 5.4 todos los valores de e son > 0.6, lo que nos habla que son distintos a un árbol de vinculación preferencial, sin embargo hay dos familias que están mas cercanas a este valor, la Familia Autronesia y la Familia Niger-Congo, para estas dos esperaríamos un comportamiento aproximado a una árbol con distribución de grado de salida de ley de potencia, lo cual abordaremos y calculamos mas adelante. La topologia de los arboles de clasificación de los lenguajes, es de tal forma que genera un numero considerabla de hojas (lenguajes) por familia de la que descienden, esto lo podemos observar comparando los nodos que se encuentran a una distancia dada del nodo raíz, ¿Cuantos de ellos son hojas? y ¿Cuantos son nodos padres?. Para esto elaboramos un programa (apendice) escrito en Mathematica para obtener la distribución de nodos por nivel y hojas por nivel obtenientod los siguientes graficos:

Familia	**n**	**h**	**e**
Afro-Asiatica	530	375	0.707547
Árbol Aleatorio	530		0.501087 ± 0.007195
Árbol Barabási	530		0.662139 ± 0.014368
Austronesia	1690	1175	0.695266
Árbol Aleatorio	1692		0.486341 ± 0.009958
Árbol Barabási	1692		0.663042 ± 0.011564
Indo-Europea	592	439	0.741554
Árbol Aleatorio	592		0.493431 ± 0.013227
Árbol Barabási	592		0.668210 ± 0.016351
Niger-Congo	1030	701	0.680582
Árbol Aleatorio	1030		0.495747 ± 0.014966
Árbol Barabási	1030		0.664570 ± 0.009364
Sino-Tibetana	563	449	0.797513
Árbol Aleatorio	563		0.505497 ± 0.008277
Árbol Barabási	563		0.669318 ± 0.012922
Trans-Neoguineana	656	477	0.727134
Árbol Aleatorio	656		0.492885 ± 0.017604
Árbol Barabási	656		0.664855 ± 0.005513

Cuadro 5.4: Se puede observar los valores de e para cada una de las familias lingüísticas. Todas las familias tiene valores superiores a 0.6. Podemos decir que para cada uno de los árboles genealógicos de los lenguajes mas del 60 % de sus nodos son hojas, lenguas actuales


Distribución de nodos en niveles
(Familia Afro−Asiatica)
Frecuencia
Nivel
Nodos
Hojas
Distribución de nodos en niveles
(Familia Austronesian)
Distribución de nodos en niveles
(Familia Indo−Europea)
Distribución de nodos en niveles
(Familia Niger−Congo)
Distribución de nodos en niveles
(Familia Sino−Tibetana)
Distribución de nodos en niveles
(Familia Trans New)

Cuadro 5.5: Se

5.3.4. Distribución de grado k

Para comenzar obtenemos la gráfica de probabilidad de las redes genealógicas de los lenguajes cuadro. Como se puede observar, la mayor parte de los nodos tienen grado $k = 1$ para todas las familias lingüísticas, y va disminuyendo el número de nodos con grados superiores. Estas distribuciones nos sugieren una distribución de potencias, sin embargo, para poder corroborar este comportamiento es necesario ajustar una función de este tipo a los datos que tenemos.

La función de distribución acumulativa se define como:

$$P_k = \sum_{k'=k}^{\infty} p_{k'}. \tag{5.2}$$

P_k es la fracción de vértices que tienen grado k o mayor. Una forma posible de interpretar esta expresión es como la probabilidad de elegir aleatoriamente un vértice de grado k o mayor. Si suponemos que el grado de la distribución sigue una ley de potencias, es decir $p_k = Ck^{-\alpha}$ para $k \geq k_{min}$ para algún k_{min}. Entonces para $k \geq k_{min}$ tenemos:

$$\begin{aligned} P_k &= C \sum_{k'=k}^{\infty} k'^{-\alpha} \simeq C \int_k^{\infty} k'^{-\alpha} dk' \\ &= \frac{C}{\alpha - 1} k^{-(\alpha-1)}, \end{aligned} \tag{5.3}$$

donde se aproximo la suma por la integral. En este punto se asume que k es lo suficientemente grande y que $\alpha > 1$ para que la integral converja. Entonces si la distribución p_k sigue una ley de potencias, entonces la función de distribución acumulativa P_k también lo hace pero con un exponente $\alpha - 1$ que es menor que el exponente original.

Las distribuciones acumulativas obtenidas para las redes lingüísticas las podemos ver en el cuadro 5.7. Aunque se podría calcular el exponente de la distribución gráficando la función de distribución acumulativa en escala $Ln - Ln$ y ajustando una recta a los puntos, la mejor practica recomendada por los investigadores en redes, es directamente de los datos, por medio de la ecuación [2]:

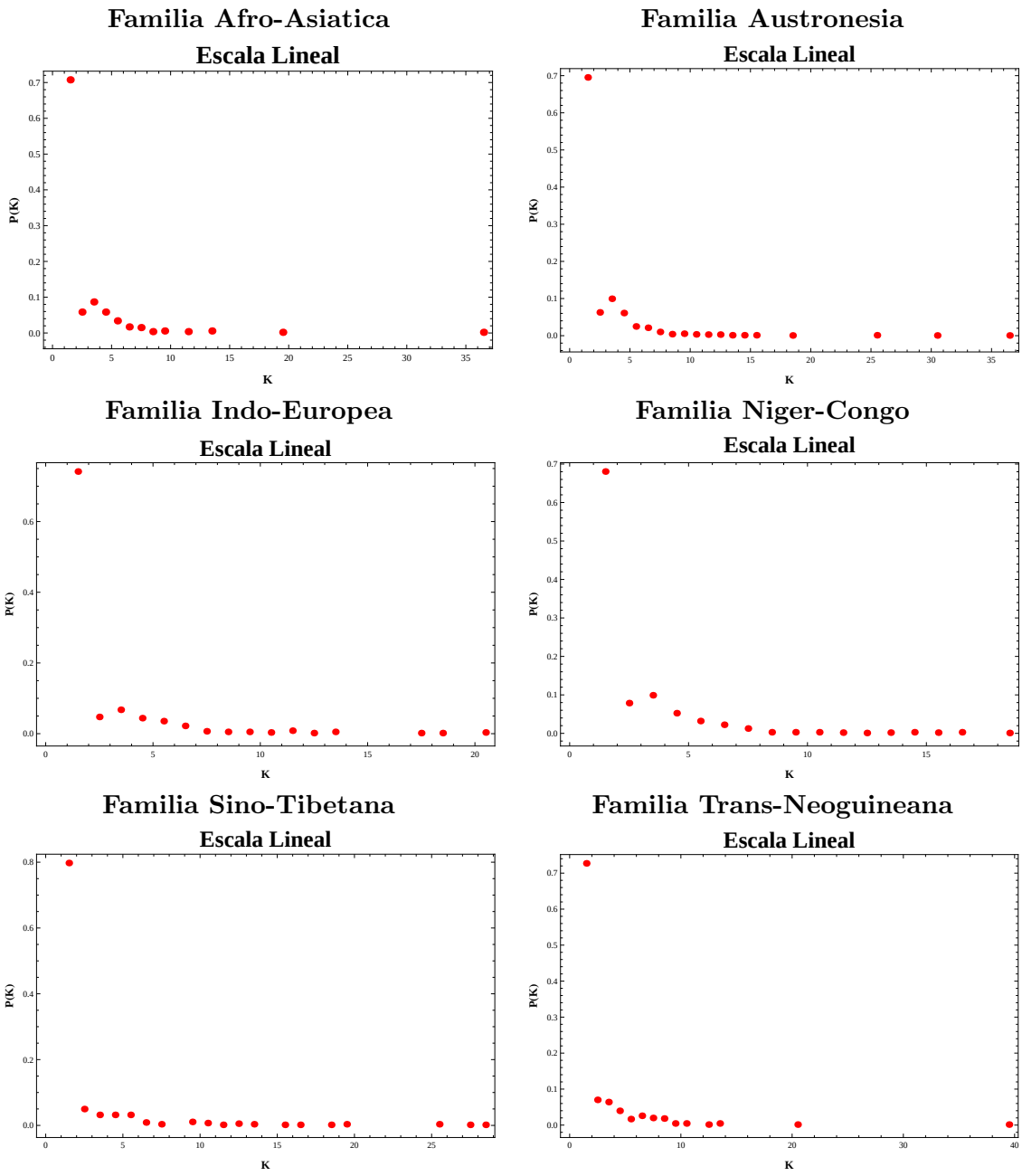
$$\alpha = 1 + N \left[\sum_i ln \frac{k_i}{k_{min} - \frac{1}{2}} \right]^{-1} \tag{5.4}$$

Donde k_{min} es el grado mínimo para el cual a partir de el cual se sigue una ley de potencias, N es el número de nodos con grado mayor o igual que K_{min}. La suma se realiza sobre todos los nodos con $k \geq k_{min}$, y no sobre todos los nodos de la red. El

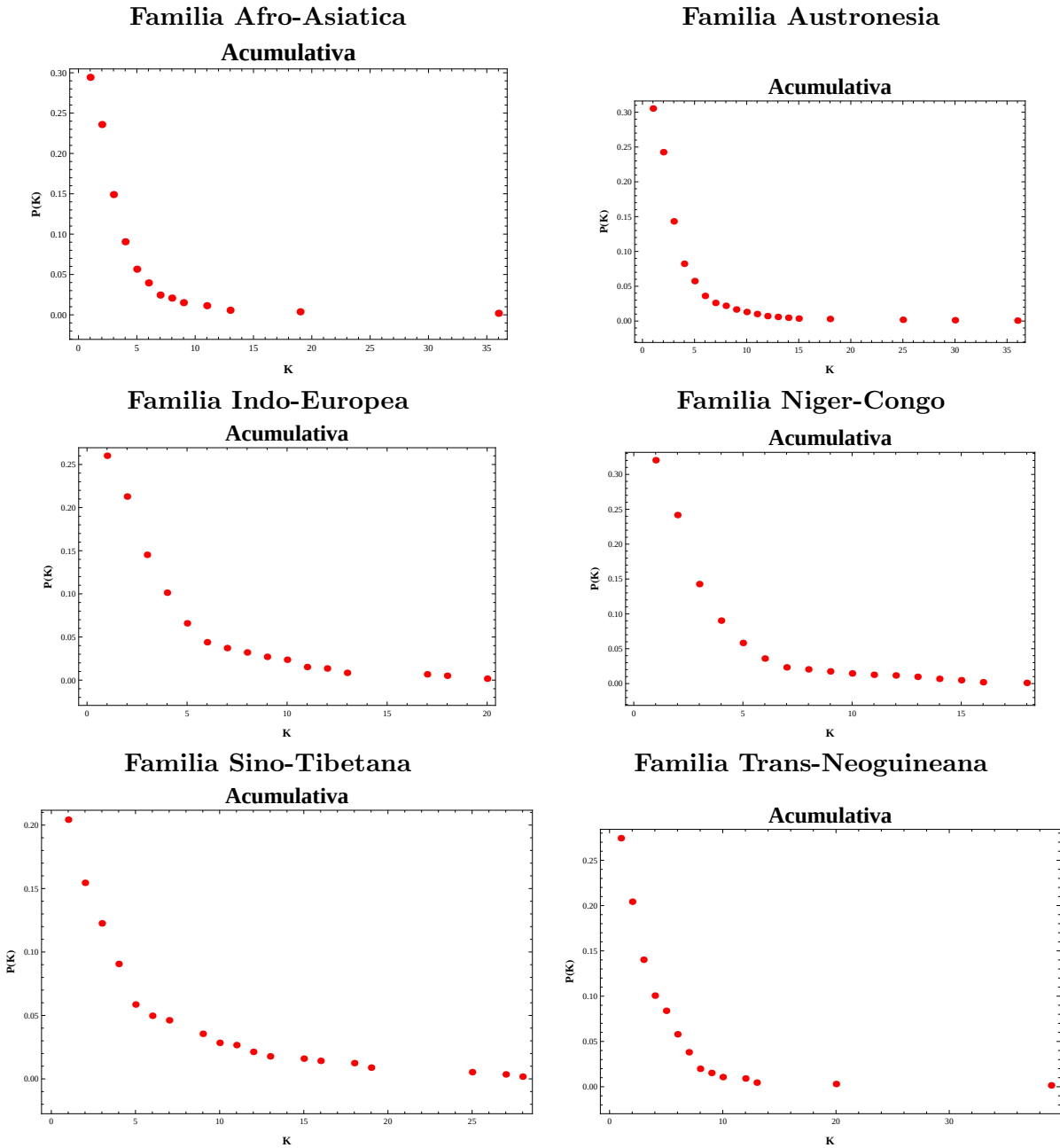
error estadístico para este calculo esta dado por:

$$\sigma = \sqrt{N}\left[\sum_i ln\frac{k_i}{k_{min}-\frac{1}{2}}\right]^{-1} = \frac{\alpha-1}{\sqrt{N}}. \tag{5.5}$$

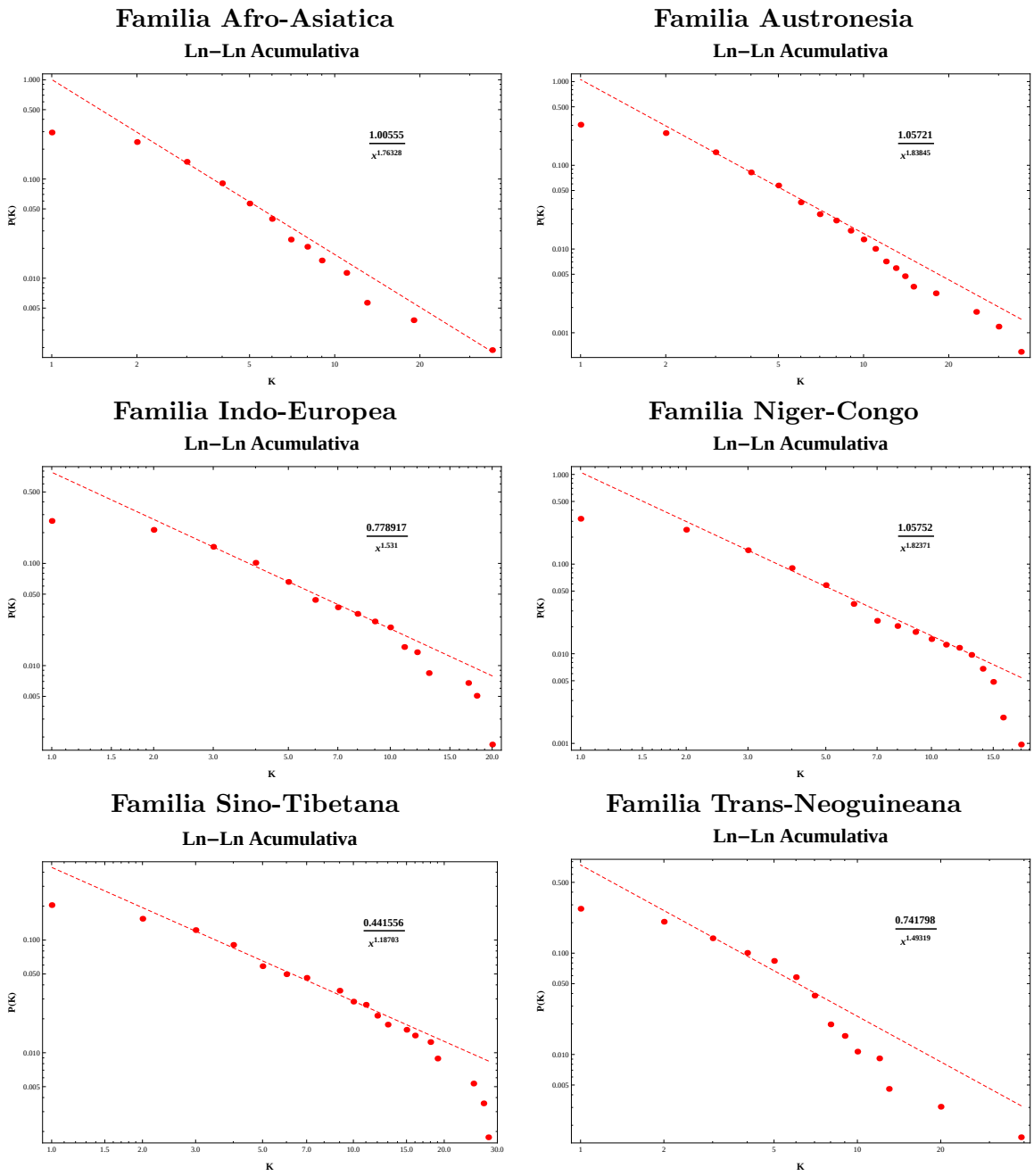
Para nuestros datos utilizamos $k_{min} = 3$ puesto que valores mas grandes provocarían que el número de datos a los cuales se les aplica el modelo fuera muy reducido. Con las ecuaciones 5.4 y 5.5 obtenemos el cuadro 5.9. Finalmente se presenta el cuadro 5.8, en este se gráfica la función de distribución acumulativa en Ln-Ln para los datos empíricos, la linea roja es la función de distribución acumulativa que mejor se ajusto a nuestros datos, también en escala logarítmica.



Cuadro 5.6: Se muestran la distribución de frecuencias del grado para las seis familias lingüísticas.



Cuadro 5.7: Distribución acumulativa de cada una de las seis familias lingüísticas.



Cuadro 5.8: Se muestra la distribución acumulativa en escala logarítmica junto con la curva de ajuste, la cual en esta escala aparece como una línea recta.

Familia	n	α	σ	r
Afro-Asiatica	530	2.763282	0.158347	0.998170
Austronesia	1690	2.838450	0.090905	0.999235
Indo-Europea	592	2.531000	0.136937	0.996981
Niger-Congo	1030	2.823707	0.115805	0.997432
Sino-Tibetana	563	2.187030	0.128001	0.995869
Trans-Neoguineana	656	2.493190	0.129476	0.981998

Cuadro 5.9: Se muestran los valores del exponente α para cada una de las distribuciones que se ajustaron a la distribución de grado de las familias lingüísticas, junto con la desviación estandar σ, así como también, el índice de correlación r entre los valores generados por el modelo y los valores empíricos

Conclusiones

Con este trabajo logramos construir un índice que determina el grado de concentración de los nodos de una red jerárquica en torno a un nivel y al cual llamamos índice de desigualdad, este valor nos permite hacer comparaciones cuantitativas con otro tipo de redes, como en nuestro caso, árboles aleatorios y arboles de tipo Barabási. En los resultados obtenemos que cuatro de las redes lingüísticas analizadas, son muy cercanas a las redes tipo Barabási, en términos del índice de desigualdad. Los dos casos que no se ajustaron a este comportamiento son, la red de la familia lingüística Austronesian, la cual presenta un índice de desigualdad, mayor incluso que una red aleatoria. Este valor resalta las limitaciones del índice propuesto, cuando comparamos redes cuya distribución de nodos en los distintos niveles se encuentran en torno a un nivel es decir, siguen una distribución de tipo campana, podemos distinguir claramente entre distintas topologías y su índice asociado. Sin embargo, cuando la red tiene una distribución bimodal como es el caso de la red Austranesian, esta clara distinción se oscurece. Para poder determinar, el comportamiento de las redes lingüísticas determinamos la distribución del grado k de cada una de estas, obteniendo un valor para el exponente que esta en el rango de redes libres de escala.

Apéndice A

Programa para generar distribución en niveles

Programa en C++.

```
////////////////////////////////////////////////////////////////////
//                  DISTRIBUCION DE N AGENTES DADOS                //
//   EN ARBOLES JERARQUICOS Y SU INDICE DE DESIGUALDAD INTERNA     //
////////////////////////////////////////////////////////////////////

#include <iostream>
#include <cmath>
#include <new>
#include <iomanip>
using namespace std;


int main()
{

int /*nodos,*/ i, j, k, l, m, n, u;
 unsigned long long int nodos;
int *agentes = new int;
unsigned long long int *nod = new unsigned long long int;
unsigned long long int *decision;
int *distrin;

//declaraciones para id.
int p, q;
double suma, dif;
double id;

//íAqu acaban las declaraciones para id.

cout << "Introduzca_el_numero_de_agentes_a_distribuir";
cin >> *agentes;
cout << "\n";
```

```
////////////////////////////////////////////////////////////////////////
// En esta parte del programa calculamos el numero de vertices que    //
//tendra nuestro arbol de decisiones, el cual es un arbol binario con //
//un nodo mas  el nodo cero                                           //
////////////////////////////////////////////////////////////////////////

        nodos = 1;

        for (i = 1; i < *agentes ; i++) /*tiene que ser n-1 para llegar */
        {

            nodos = nodos + pow(2,(i-1));
        }
        //cout << " El numero de nodos para " << *agentes << "agentes es "
             << nodos << endl;


////////////////////////////////////////////////////////////////////////
//Despues de calcular el numero de vertices que tendra nuestro arbol//
//de decisiones para n agentes (o niveles). Se construye el arbol.  //
//En cada paso de nuestro arbol binario solo se puede tomar el mismo//
//valor o una unidad mayor que el padre.                             //
////////////////////////////////////////////////////////////////////////

        *nod = nodos;                          // Numero de entradas del
            arreglo
        decision = new unsigned long long int[*nod]; // Arreglo con
            entradas dinamicas



        decision[0] = 0;
        decision[1] = 1;
        for ( j = 1 ; j < (*nod / 2); j++)              // *nod/2 puesto
            que en cada pasada son 2 elementos.
        {
           decision[ 2 * j ] = decision[j];             // El hijo par
               toma el valor del padre.
           decision[ 2 * j + 1] = decision[j] + 1;      // El hijo impar
               toma el valor del padre mas 1.
        }

 ////////////////////////////////////////////////
 //Mostramos en pantalla el arreglo de decisiones //
 //  resultante para un numero dado de agentes.   //
 ////////////////////////////////////////////////

//       for ( k = 0; k < *nod; k++)
//       {
//          cout << decision[k] << " ";
//       }
//       cout << "\n" << endl ;
```

```
////////////////////////////////////////////////////
//Formamos el path-arreglo. Que es la distribucion   //
// de los n agentes en los distintos niveles.        //
   ////////////////////////////////////////////////////



        distrin = new int[*agentes];
 for ( u = 0; u < pow(2,(*agentes -2)); u++)
 {
        l = pow(2,(*agentes -2)) + u ;

        distrin[*agentes - 1] = decision[l];

        for ( m = (*agentes-2); m >= 0; m--)
        {
            if ( l %2 == 0)
            {
               distrin[m] = decision[l / 2];
               l = l / 2;
            }
            else
            {
               distrin[m] = decision[( l - 1) / 2];
               l = (l - 1) / 2;
            }
         }
         cout << setw(9/*pow(2,(*agentes -2))*/ + 1 )        //Falta
             averiguar como paso de un numero al numero de digitos
             << setiosflags(ios::left) << u + 1
             << setw(10)                                    // Lo mejor es
                  poner 4 puesto que no pasamos de cifras de 3 digitos
             << setiosflags(ios::left) << *agentes;

        for ( n = 0; n < *agentes; n++)
        {

           cout << setw(3) << setiosflags(ios::left) << distrin[n];
        }

 ////////////////////////////////////////////////
 // Calculamos el ID para el arreglo distrin.   //
    ////////////////////////////////////////////////

        suma = 0;

        for ( p = 0; p < *agentes; p++)
        {
           for ( q = 0; q < *agentes; q++)
           {
                dif = abs(static_cast<double>(distrin[p] - distrin[q]));
                suma = suma + dif;
           }
```

```
        }

        id = suma / (2*(*agentes)*(*agentes - 1) );

        cout << "    "
             << setw(7) << fixed
             << setprecision(4)
             << setiosflags(ios::left)
             << id;




      cout << endl;
 }


////////////////////////////////////////////////////////////
//Finalmente liberamos la memoria que estamos utilizando.  //
//  Es decir mandamos a la pila la memoria establecida de  //
//              manera dinamica nuevamente.                //
////////////////////////////////////////////////////////////

    delete[] agentes;
    delete[] nod;
    delete[] decision;
    delete[] distrin;


return 0;
}
```

Apéndice B

Programa para generar arboles aleatorios

Programa en Mathematica 9.

```
arbalemedidaslen = CreateDirectory[“arbalemedidaslen”];
SetDirectory[arbalemedidaslen];
Do[
medidas =
Table[
arbale = TreeGraph[RandomInteger[#] → # + 1&/@Range[0, i − 2]];
n = VertexCount[arbale]; dis = Sort[GraphDistance[arbale, 0]];
disprom = Mean[dis]//N; dmax = Max[dis]; Ides = Id[dis];
LiOci = {First[#], Length[#]}&/@Split[dis]; Iext = Ie[LiOci];
hojas =
Length[
Select[
Sort[Partition[Riffle[VertexList[arbale],
VertexOutDegree[arbale]], 2], #1[[2]] < #2[[2]]&],
#[[2]] == 0&]]; {n, disprom, dmax, Ides, Iext, hojas}, {3000}];
Export[ToString[n] <> “.dat”, medidas]; ,
{i,3,2000,10
```

```
];
```

Apéndice C

Programa para generar arboles con vinculación preferencial

Bibliografía

[1] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[2] Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.

[3] Joseph Persky. Retrospectives: Pareto's law. *The Journal of Economic Perspectives*, pages 181–192, 1992.

[4] Sitabhra Sinha and Nisheeth Srivastava. Is inequality inevitable in society? income distribution as a consequence of resource flow in hierarchical organizations. In *Econophysics of Markets and Business Networks*, pages 216–226. Springer, 2007.

[5] Joseph E Stiglitz and Alejandro Pradera. *El precio de la desigualdad*. Punto de Lectura, 2014.

[6] Steven Strogatz. *Sync: The emerging science of spontaneous order*. Hyperion, 2003.

[7] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442, 1998.