FAMAF Facultad de Matemática, Astronomía, Física y Computación

UNC Universidad Nacional de Córdoba

Comunidades y robustez en redes complejas reales y sintéticas

por

Nahuel Almeira

Presentado ante la Facultad de Matemática, Astronomía, Física y Computación como parte de los requerimientos para la obtención del grado de Doctor en Física de la

UNIVERSIDAD NACIONAL DE CÓRDOBA

Agosto, 2022

Director

Dr. Orlando Vito Billoni

Tribunal especial (titulares y suplentes)

Dr. Marcelo KUPERMAN (Instituto Balseiro)
Dra. Silvia Adriana MENCHON (FaMAF)
Dr. Martín Ariel DOMINGUEZ (FaMAF)
Dra. María Fabiana LAGUNA (Centro Atómico Bariloche)
Dr. Jorge Alberto REVELLI (FaMAF)
Dr. Rodolfo Guillermo PEREYRA (FaMAF)

Comunidades y robustez en redes complejas reales y sintéticas

Nahuel Almeira

Resumen

El estudio de las redes complejas –sistemas cuya estructura está regida por una compleja red de interacciones– es un área activa de investigación multidisciplinaria en la cual la física ha tenido un papel preponderante. Este campo ha recibido un gran impulso en los últimos años debido a la disponibilidad de una cantidad masiva de datos y a la existencia de recursos computacionales que permiten llevar a cabo los análisis estadísticos requeridos. Redes complejas extensas pueden encontrarse en una gran cantidad de sistemas naturales y artificiales, tales como sistemas físicos, biológicos, sociales, infraestructuras tecnológicas, etc. Dada la importancia de estos sistemas, las investigaciones que exploran la estructura, la dinámica y la evolución de las redes complejas ha concitado el interés de la comunidad de la física, ya que las herramientas de la mecánica estadística, así como las de otros campos de esta disciplina, tienen una aplicación directa para analizar y entender estos sistemas. La caracterización estadística de la estructura de las redes complejas se aborda desde diversos enfoques. Entre ellos hay dos que se destacan por la información que aportan y por las implicancias prácticas que ofrecen. Por una parte, la caracterización de estructuras modulares o comunidades es importante para entender la funcionalidad de las redes. Por otra parte, el estudio de la resiliencia de las redes ante fallas o ataques dirigidos es de gran utilidad para comprender cómo pueden generarse redes con un funcionamiento robusto. Curiosamente, estos dos conceptos –la existencia de comunidades y robustez de las redes– están íntimamente relacionados, y su estudio presenta grandes desafíos, dada la complejidad de los cálculos y análisis necesarios para su abordaje. En esta tesis estudiamos la existencia de comunidades en redes de jugadores de ajedrez utilizando la base de datos de partidas más extensa disponible, en su momento, en el mundo. Realizamos una caracterización general de las mismas y observamos una fuerte correlación entre las comunidades y el nivel de juego de los jugadores. En lo que respecta a las fallas y/o ataques analizamos redes sintéticas, tales como grafos de Erdös-Rényi y redes planares de Delaunay. Caracterizamos la robustez de las mismas mediante el estudio de transiciones de percolación, utilizando las herramientas de análisis de los fenómenos críticos y extensas simulaciones numéricas. Observamos que las transiciones varían de manera cualitativa de acuerdo con el tipo de red y con la estrategia de ataque empleada. En particular, observamos que algunos ataques generan transiciones similares a las encontradas en procesos de percolación explosiva.

Communities and robustness of real and synthetic complex networks

Nahuel Almeira

Abstract

The study of complex networks –systems whose structure is governed by a complex interaction network– is an active multidisciplinary field of research in which physics has had a prevailing role. This field has received great impulse during the last years because of the availability of a massive amount of data and the existence of computational resources that allow to perform the required statistical analysis. Extensive complex networks can be found in a large variety of natural and artificial systems, such as physical, biological, and social systems, technological infrastructures, etc. Given the importance of these systems, investigations exploring the structure, dynamics and evolution of complex networks has raised the interest of the physics community, as tools coming from statistical mechanics, as well as from other fields of physics, are fundamental and have a direct application in the analysis and comprehension of such systems. The statistical characterization of the structure of complex networks is addressed from different approaches. Among them there are two that stand because of the information they give and because of the practical applications they allow. On one side, the characterization of modular structures or communities is important to understand the functionality of networks. On the other side, the study of the resiliency of networks against failures or targeted attacks gives relevant information on how to develop robust networks. Interestingly, these two concepts –the existence of communities and the robustness of networks– are closely related, and their study presents big challenges, given the complexity of the calculations and analysis required to address them. In this thesis we studied the existence of communities in networks of chess players using the largest available database, at the moment, on the world. We performed a general characterization of the networks, observing a strong correlation between communities and player skill level. In terms of failures and targeted attacks, we studied synthetic random networks, such as Erdös-Rényi graphs and Delaunay triangulations. We characterized their robustness through the study of percolation transitions, using tools from critical phenomena and extensive numerical simulations. We observed that the transitions vary in a qualitative manner depending on the type of network and the attack strategy. In particular, we found that certain attacks generate transitions similar to those encountered in processes of explosive percolation.

Dedicatoria

A mi hijo Alfonso.

Agradecimientos

A mi familia, que estuvo, está y estará siempre a mi lado. En especial a mi hermana Cande y a mis viejos, Clau y Ger.

A mi gran amor Mica, por darme fuerzas y brindarme siempre su apoyo constante.

A mi director Orlando, que con la generosidad y bondad que lo caracteriza supo motivarme y guiarme durante todo este proceso.

A mi colega Juan, quien ofició informalmente como co-director, siempre aportando nuevos puntos de vista y discutiendo apasionadamente cada idea.

A Ana y Andrés, con quienes tuve la suerte de colaborar y de quienes aprendí mucho.

A mis amigos, en especial a Manu y Becu, con quienes disfruté del reencuentro en Córdoba.

A "los poetas", que me demostraron que la amistad no sabe de distancias.

A la FaMAF y a CONICET, instituciones que hicieron posible este proyecto.

Índice general

Capítulo 1

Introducción

El campo de las redes complejas es una disciplina bastante reciente que ha influenciado fuertemente a varias áreas del conocimiento tales como la física estadística, la informática, la biología y la sociología. Su desarrollo comienza aproximadamente a inicios de este siglo [AB02; New18], inspirado en hallazgos empíricos en redes asociadas a las disciplinas antes mencionadas. Por otra parte, la disciplina se ha visto impulsada por la existencia de grandes volúmenes de datos y por el crecimiento sostenido en el poder de cómputo de los actuales equipos informáticos que permiten generar y analizar redes complejas extensas.

En contraste con otras áreas de investigación, muchos de los problemas asociados a las redes complejas son fáciles de definir en base a situaciones que podrían ocurrir en la vida cotidiana. Por ejemplo, podríamos intentar responder el interrogante de cómo detener una epidemia utilizando la mínima cantidad de vacunas si se conoce la red de contactos entre las personas [Alv+19], o si es posible identificar a grupos de personas que son socialmente afines en una red social o en una comunidad [Zac77], o también si se puede hacer que las redes de transmisión eléctricas sean robustas ante fallas o ataques intencionales, de modo que se puedan evitar los apagones [Bul+10]. Para resolver estas preguntas, la disciplina de las redes complejas ha desarrollado varias herramientas y métricas que permiten caracterizar una diversidad de redes empíricas tomadas de distintos ámbitos así como redes sintéticas generadas por distintos modelos. De esta manera se busca comprender, en forma cada vez más precisa, el funcionamiento de los procesos que se desarrollan sobre estas redes.

En términos generales las redes complejas están compuestas por unidades o nodos que interaccionan entre sí por medio de conexiones o enlaces, formando diversas estructuras. La complejidad radica en que la distribución de interacciones combina un cierto grado de orden con una importante componente aleatoria, de manera que la caracterización de las redes no es sencilla y muchas de las herramientas desarrolladas para su estudio tienen un carácter estadístico.

Podemos dividir el trabajo realizado en esta tesis en dos partes. Una de las partes se refiere a la caracterización de una red empírica de jugadores de ajedrez, donde nos interesamos particularmente en la relación entre comunidades y atri-

butos de los nodos [Alm+17]. En la otra parte, analizamos dos modelos de redes sintéticas, con características bien diferenciadas. En este caso, profundizamos los estudios existentes respecto a la robustez de las redes ante ataques dirigidos empleando las herramientas de los procesos de percolación [ABP20; Alm+21].

Se conoce que en una red social existen personas que, por sus intereses en común u otras afinidades, tienden a vincularse estrechamente formando grupos o comunidades. Este hecho empírico genera el interrogante de si se pueden identificar a las comunidades conociendo únicamente las relaciones entre las personas, sin conocer la naturaleza del vínculo que las une, y si estas comunidades además guardan relación con algún aspecto particular de las personas que las componen. En forma abstracta se puede definir a la detección de comunidades como el problema de identificar conjuntos de nodos que se encuentran densamente conectados entre sí y que a su vez están menos conectados con los nodos de otras comunidades. Si bien este es uno de los muchos aspectos que se pueden estudiar en una red compleja, la identificación de comunidades es uno de los más relevantes. La detección de comunidades ofrece información clave respecto de cómo está formada la red y permite enfocarse en regiones de la red que tienen un cierto grado de autonomía. Además, se pueden clasificar a los nodos respecto del rol que cumplen dentro de la comunidad a la cual pertenecen. Por ejemplo, la función que cumplen los nodos que están completamente embebidos en una comunidad es diferente a la de aquellos nodos que sirven de enlace o puente entre las comunidades. Estos últimos mantienen a las comunidades unidas y además juegan un papel más importante en la propagación de información o el control de procesos dinámicos que se desarrollan sobre las redes [Gra73].

El problema de la detección de comunidades, si bien es un problema aún abierto respecto al objeto que se busca identificar, ha generado en los últimos años el desarrollo de muchos algoritmos que posibilitan encontrar comunidades de modo eficiente y cuyo funcionamiento se basa en distintos principios [FH16]. Estos algoritmos pueden clasificarse de acuerdo a la forma en que operan o el concepto de comunidad que utilizan. Como se trata de un problema abierto, no existe un procedimiento directo o consensuado para comparar la efectividad de los diferentes métodos. Sin embargo, en la mayoría de las aplicaciones prácticas se emplea un número reducido de algoritmos. En este sentido, uno de los métodos más empleados debido a su rapidez y efectividad es el algoritmo de Louvain [Blo+08], el cual también se puede usar en redes con pesos en sus conexiones. Otros métodos para la detección de comunidades utilizan información adicional sobre los nodos (atributos) que no está necesariamente codificada en las interacciones o conexiones entre los nodos. En este sentido, Newman y Clauset [NC16] introdujeron un algoritmo para la detección de comunidades que hace uso de la inferencia bayesiana, mediante la cual se construye un modelo generativo de redes. En esta aproximación al problema las redes generadas emulan algunas de las características de las redes que se quiere estudiar, tales como la estructura de comunidades y atributos de los

nodos. Utilizando este tipo de modelos es posible estudiar correlaciones entre la topología de la red (su estructura de conexiones) y los atributos asociados a los nodos.

Dado que nuestro grupo contaba con bastante experiencia en el análisis de correlaciones en bases de datos de partidas de ajedrez [Per+13; SPB14; SPB16] nos propusimos extender estos estudios al campo de las redes complejas. Ya de por sí, resulta interesante caracterizar a la red de jugadores de ajedrez utilizando distintas métricas, a fin de realizar un análisis comparativo con otras redes sociales y tecnológicas. Además, el estudio de correlaciones entre la topología de una red social y los atributos de sus nodos es un tópico importante y bastante actual en el área de las redes complejas [Leo+16; NC16; HPF16; HDF14]. En las redes de ajedrez se pueden asociar distintos atributos a los nodos. En particular, existe una métrica llamada Elo [Gli95], cuyo valor numérico representa el nivel de juego de cada jugador. Mas allá de la bases de datos en particular que nos propusimos analizar, la importancia del estudio radica en que se puede obtener información valiosa para entender la estructura y la dinámica de las comunidades sociales.

Nuestro estudio sobre la caracterización de las redes de jugadores de ajedrez consideró dos grandes bases de datos. Una que registra partidas jugadas sobre el tablero (OTB) y otra que considera juegos virtuales desarrollados en portales de internet (Portal). Encontramos que estas redes tienen modos particulares de evolución que las lleva a presentar diferencias y similitudes. En ambos casos las distribuciones de grado son heterogéneas y en el caso de Portal se aproxima bastante bien a una ley de potencias. Una de las principales diferencias entre las redes se observa en las correlaciones entre los grados de los nodos. OTB es una red asortativa, como es el caso de muchas redes sociales [New02], en cambio Portal tiene una estructura disortativa. Observamos que en las dos redes, OTB y Portal, los jugadores de élite tienden a formar grupos densamente conectados, esto se ve reflejado en el coeficiente de *club de ricos* normalizado que introdujimos para analizar el efecto. Por último, empleando diferentes algoritmos para al detección de comunidades confirmamos que existe una correlación significativa entre la estructura de comunidades y el desempeño de los jugadores (su nivel de Elo).

Otro aspecto muy importante en la caracterización de las redes complejas es el que se refiere al estudio de la robustez o resiliencia de las mismas ante ataques o fallas. Muchas veces los sistemas asociados a estas redes pierden uno o varios de sus componentes (nodos y/o conexiones) ya sea por una falla o por un ataque intencional. Surge entonces la pregunta de cómo se ve afectado el funcionamiento del sistema en relación a su dinámica cuando la estructura o topología de la red se ve afectada. Expresado de otra manera, es de interés conocer cuáles son los aspectos estructurales que confieren estabilidad a las redes y a las dinámicas que operan sobre las mismas. Estos problemas se estudian analizando una dinámica en particular o desde una perspectiva más general considerando únicamente la estructura o topología de la red. En ambos casos

interesa conocer cuál es la fracción de conexiones o nodos que tiene que fallar para que el sistema deje de funcionar en su conjunto. Cuando se considera el problema desde el punto de vista estructural, el interrogante se puede plantear desde dos puntos de vista. Uno de ellos se refiere a saber qué nodos o conexiones conviene atacar primero para producir el mayor daño en la red; en este sentido, el objetivo final es fragmentar la red de forma eficiente. El otro punto de vista consiste en considerar la protección del sistema. La cuestión entonces es saber cuáles nodos deben protegerse para evitar fallas en la funcionalidad o, dicho de otra manera, qué estructura debe adoptar una red para ser menos vulnerable. En este último enfoque el objetivo es diseñar un buen mecanismo de defensa ante ataques o de resiliencia ante fallas.

El problema de la fragmentación de redes ante ataques o fallas se conecta con el estudio de formas eficientes para el desmantelamiento de redes [Bra+16]. Por otra parte, cuando se busca comparar los mecanismos de ataques y fallas de forma rigurosa el marco conceptual adecuado es el de los procesos de percolación [SA18]. La fallas son usualmente modeladas como la remoción aleatoria de nodos o conexiones, mientras que en los ataques se remueven los nodos o conexiones mas influyentes. Para ello se utiliza una lista que los ordena de acuerdo a su importancia, siguiendo algún criterio preestablecido. De esta manera se intenta producir el mayor daño a la red removiendo la menor cantidad de nodos o conexiones.

Un criterio para determinar la importancia de los nodos o de las conexiones ante ataques o fallas consiste en considerar medidas de centralidad [Iye+13]. En este tipo de ataques se establece un orden de prioridad en el que se eliminan primero los nodos o las conexiones más centrales. Entre las medidas de centralidad para los nodos más utilizadas se encuentran el grado (cantidad de vecinos cercanos), y el *betweenness* [Fre77]. Esta ultima medida es global, y representa la fracción de los caminos geodésicos entre pares de nodos de la red que pasan por un dado nodo. Las dos medidas miden la carga o información que circula sobre los nodos; una localmente y la otra globalmente. En un trabajo bastante reciente, Wandelt et al. [Wan+18] realizaron un estudio comparativo de la efectividad de una variedad de estrategias de desmantelamiento. Para esto emplearon un conjunto de redes de prueba, incluyendo redes sintéticas y reales. Este estudio comparativo es el mas extenso que existe a la fecha, e incluye tanto algoritmos de desmantelamiento como ataques basados en centralidad. Como regla general, la efectividad de la estrategia de ataque depende del tipo de ataque utilizado y de la topología de la red.

En un marco general, es importante conocer qué estructuras de red son robustas ante distintos tipos de ataques o fallas, o cómo pequeñas modificaciones pueden mejorar la resiliencia de una red [Sch+11; ZL12], preservando a la vez ciertas características como, por ejemplo, su estructura de comunidades [Yan+15]. Además, los conceptos generales que surgen del análisis de estos problemas tienen un alto valor ya que permiten entender el funcionamiento de un amplio conjunto de sistemas. Es sabido que las redes heterogéneas son

frágiles ante ataques dirigidos por medidas de centralidad [Cal+00; AJB00; Coh+00; Hol+02], mientras que son robustas ante fallas aleatorias [Coh+00]. Por otra parte, redes con una distribución homogénea de grado, tales como los grafos de Erdős-Rényi (ER), se espera que sean robustas ante ataques dirigidos. En particular se sabe que son robustas ante ataques que utilizan los grados de los nodos como medidas de centralidad [Cru+04]. Sin embargo, algunas redes son frágiles ante ataques dirigidos a pesar de tener una distribución de grado homogénea. Uno de estos ejemplos es la red de transmisión eléctrica de los Estados Unidos, la cual muestra una perdida significativa de la conectividad cuando los nodos con alta carga son removidos [AAN04]. Otro ejemplo es el modelo de Watts-Strogatz para redes de mundo pequeño. Estas redes son particularmente frágiles en un escenario de fallas en cascada, como mostraron Xia et al. [XFH10]. Estos autores atribuyeron la fragilidad de estas redes a que poseen una distribución de betweenness heterogénea. Por otra parte se sabe que los ataques por betweenness clasifican entre los mas eficientes para desmantelar o fragmentar una red [Hol+02; CG17; Iye+13; Wan+18].

El modelo de Erdős-Rényi (ER) se utiliza para la generación de redes con un alto grado de desorden. En estas redes, cada par de nodos se conecta al azar con una dada probabilidad, que es la misma para todos los pares de nodos, dando como resultado una red que no tiene correlaciones entre nodos. Además, tanto la distribución de grado como la de betweenness son homogéneas [XFH10; Kor+18]. Este modelo tiene una gran importancia ya que es un paradigma para analizar redes empíricas homogéneas, sin embargo cuando las redes muestran correlaciones entre patrones no resulta un modelo adecuado. En particular, muchos sistemas complejos con distribuciones homogéneas se encuentran restringidos por vínculos espaciales. Por ejemplo, las redes de transmisión eléctrica, las redes de movilidad, Internet, el cerebro, etc. Todas estas redes evolucionan bajo la influencia de aspectos geométricos. En estos sistemas es conveniente reemplazar el modelo de ER por modelos de redes espaciales en dos y tres dimensiones, en donde los nodos y las conexiones están embebidos en un espacio [Bar11].

Entre los modelos realistas para estudiar las redes embebidas se encuentran los modelos espaciales aleatorios. Uno de los más estudiados en la literatura de los sistemas físicos es la teselación de Voronoi y su dual, denominado triangulación de Delaunay (DT) o entramado aleatorio (*random lattice*) [BZ09; Kir+18]. Es importante mencionar que ciertas métricas locales que se utilizan de manera corriente en la caracterización de las muchas redes, tales como el grado de los nodos, la transitividad, la asortatividad, etc., no ofrecen mucha información en el caso de las redes espaciales. Por ejemplo, una red espacial planar usualmente tiene una distribución de grado centrada. Por otro lado, en una triangulación la transitividad se maximiza trivialmente. En cambio, otras medidas globales tales como las distancias topológicas y geométricas se vuelven más relevantes. En conclusión, en los modelos de redes planares como DT se puede esperar que exista un marcado contraste entre las métricas locales y

globales.

Tanto el modelo de ER como el DT son un punto de partida interesante para estudiar la robustez ante ataques dirigidos en redes complejas utilizando medidas de centralidad locales y globales. Por ejemplo, se puede analizar la validez de que en redes con distribución homogénea de grado los ataques por betweenness no son más efectivos que otros ataques dirigidos. Además, como se trata de dos modelos para la generación de redes homogéneas con características bien diferenciadas, el efecto de distintos tipos de ataques ofrece información complementaria. Por otra parte, cada una de estas dos redes sintéticas representa a un gran conjunto redes empíricas y los ataques pueden interpretarse en relación a los procesos asociados a esas redes. Por ejemplo, la conectividad es una medida particularmente conveniente para analizar la robustez en redes planares ya que una interrupción en la conectividad estructural esta seguida frecuentemente por una falla del sistema [Bul+10].

Existe en la literatura una extensa lista de trabajos donde se estudia la robustez de las redes de Delaunay, y otras redes espaciales, en el marco de la teoría de percolación [SE64; BR06; BZ09; Mel13; NMH16; Nor16; Oli+08; Din+14]. Los umbrales de percolación para la percolación por nodos (sitios) y por conexiones (enlaces) son bien conocidos [SE64; BR06; BZ09] y existe evidencia de que ambos procesos pertenecen a la misma clase de universalidad que la percolación aleatoria en redes regulares de la misma dimensión [HH99; McC87]. Además de la percolación aleatoria, las redes DT también han sido estudiadas en el contexto de ataques dirigidos basados en medidas de centralidad. Por ejemplo, dos estrategias de ataques son analizadas en [NMH16]. En uno de los ataques los nodos con el grado mas alto son removidos secuencialmente. En este caso las redes se fragmentan removiendo una menor cantidad de nodos que en el caso aleatorio, aunque la naturaleza de la transición no cambia. Es decir, los dos procedimientos pertenecen a la misma clase de universalidad en el proceso de percolación. En otra estrategia de ataque, basada en el betweenness de los nodos, las redes se fragmentan aún más rápido. En este caso el umbral de percolación en término de la fracción de nodos removidos tiende a cero en el límite termodinámico. Sin embargo, en el trabajo citado los autores no pudieron determinar la clase de universalidad de la transición.

En el análisis de los procesos de percolación estudiamos las redes de Erdős-Rényi y en la triangulación de Delaunay utilizando distintas estrategias de ataques. Para las redes de ER utilizamos versiones de ataques, iniciales y recalculados, que están basados en medidas de centralidad locales y globales. En los ataques iniciales la centralidad de los nodos a remover se calcula antes de comenzar el ataque, y la lista de nodos se mantiene durante el proceso de remoción de nodos. En el caso recalculado, la lista se actualiza después de cada remoción, calculando nuevamente las medidas de centralidad. En las redes de Delaunay, si bien analizamos varios ataques, nos enfocamos particularmente en ataques iniciales basados en betweenness, utilizando también aproximaciones al betweenness que emplean un rango acotado de información. Empleamos análisis

de escala de tamaño finito para evaluar la naturaleza de las transiciones hacia el estado fragmentado y eventualmente determinar la clase de universalidad de la transición.

En el caso de las redes ER nuestros resultados muestran que la transición inducida por el ataque por betweenness recalculado es muy abrupta, en comparación con el resto de los ataques y se desvía considerablemente de la clase de universalidad de un proceso de percolación aleatorio. Dado que la variación del parámetro de orden es muy abrupta en este ataque se puede considerar al proceso como un caso de percolación explosiva [ADS09; Da +14; DSo+19]. Los resultados del análisis de tamaño finito son consistentes con una transición de fase continua aunque no pudimos determinar que este resultado sea válido en el límite de tamaños infinitos. Por otra parte, mas allá de que las redes de ER tienen una distribución homogénea de grado y de betweenness, los ataques por betweenness superan a los ataques guiados por otras medidas de centralidad, por ejemplo el ataque por grados. En particular, la versión del ataque por betweenness que actualiza la lista de nodos a remover en cada paso es particularmente efectivo para destruir redes de ER, con una desempeño comparable a los métodos más eficientes para desmantelar redes [Bra+16; MM15].

En las redes planares construidas a partir de la triangulación de Delaunay mostramos que la transición de percolación en un ataque inicial por betweenness ocurre cuando una fracción subextensiva de los nodos es removida, tal como ocurre en la versión recalculada que fue reportada en [NMH16]. Llamativamente, la transición es compatible con una transición de primer orden lo que permite analizar este proceso—como en el caso de las redes de ER con ataques recalculados—dentro del marco de la percolación explosiva [ADS09; Da +14; DSo+19]. Respecto de los ataques basados en aproximaciones al betweenness, es decir donde sólo se consideran caminos de un alcance limitado, pudimos observar que las transiciones son continuas y muestran un umbral de percolación donde el parámetro de control es distinto de cero.

Capítulo 2

Marco Teórico

2.1. Métricas de redes

En términos matemáticos, una red compleja no es más que un grafo. Si podemos definir ese grafo, es decir, si contamos con la lista de nodos y sus conexiones, tenemos en principio todo lo que se podría saber de la red. En la práctica, no obstante, la información cruda del grafo no es fácilmente asimilable para los seres humanos. Si se tratase de una red pequeña, podríamos hacer uso de alguna visualización útil e intentar extraer información cualitativa inspeccionándola, pero este enfoque ya no funciona en redes grandes. Una alternativa mejor es definir un conjunto de medidas matemáticas que capturen características interesantes de la red de manera cuantitativa, reduciendo grandes volúmenes de datos en algunos valores que puedan ser interpretados con mayor facilidad. En los últimos años se ha propuesto una enorme cantidad de medidas y en esta sección describiremos algunas de ellas. Veremos que muchas de las ideas que discutiremos provienen de las ciencias sociales, particularmente del área de *análisis de redes sociales*. Sin embargo, con el tiempo se han ido incorporando muchas otras disciplinas, entre ellas la física, logrando grandes aportes al área.

2.1.1. Distribución de grado

Se denomina *grado* al número de conexiones de un nodo. En otras palabras, se dice que un nodo u tiene grado k si dicho nodo está conectado con otros k nodos de la red. La *distribución de grado* de una red es la distribución de probabilidad asociada al grado. Usaremos la notación p_k para indicar la probabilidad de que un nodo escogido uniformemente al azar tenga grado k. Evidentemente, p_k cumple con la propiedad de normalización

$$\sum_{k=0}^{\infty} p_k = 1. \tag{2.1}$$

Además, podemos definir a partir de la distribución de grado otras cantidades tales como el grado medio

$$\langle k \rangle = \sum_{k=0}^{\infty} k p_k. \tag{2.2}$$

Si bien es un concepto simple, la distribución de grado es un observable de extremada utilidad en el estudio de las redes complejas. Por ejemplo, es sabido que muchas redes empíricas exhiben una distribución de grado heterogénea, que puede modelarse dentro de ciertas aproximaciones como una ley de potencias $p_k \sim k^{-\gamma}$, con $\gamma > 1$. En términos coloquiales, esto quiere decir que la mayor parte de los nodos de una red posee muy pocos vecinos pero que existen algunos nodos, comúnmente llamados *hubs*, que concentran muchas conexiones.

Podemos pensar en p_k como la probabilidad de que un nodo elegido uniformemente al azar tenga grado k. Si en lugar de elegir un nodo elegimos uniformemente al azar un enlace y lo seguimos hasta alguno de sus extremos, la probabilidad de que ese nodo esté conectado con otros k vecinos es

$$q_k = \frac{(k+1)p_{k+1}}{\sum_{k=0}^{\infty}(k+1)p_{k+1}} = \frac{(k+1)p_{k+1}}{\sum_{k=0}^{\infty} k p_k} = \frac{(k+1)p_{k+1}}{\langle k \rangle}. \tag{2.3}$$

Esta distribución de probabilidad se conoce como *distribución de grado en exceso* y resulta útil en numerosos cálculos [New18].

2.1.2. Transitividad y coeficiente de *clustering*

Un aspecto particularmente importante de las redes sociales, y que resulta útil hasta cierto punto en otros tipos de redes, es el concepto de *transitividad*. En matemática, una relación "$\circ$" es transitiva si $a \circ b$ y $b \circ c$ implica $a \circ c$. El ejemplo más simple es la relación de igualdad. Si $a = b$ y $b = c$, sabemos que $a = c$, por lo que la igualdad es una relación transitiva.

La relación más fundamental que existe entre nodos de una red es la "conectividad por enlace". Si esta relación fuese transitiva, significaría que si el nodo u está conectado con v, y a su vez v se conecta con w, entonces u se conecta con w. O, dicho de manera coloquial, "el amigo de mi amigo es también mi amigo". Es este el concepto que tenemos en mente cuando hablamos de transitividad en una red.

Aunque la transitividad estricta no suele ocurrir en redes empíricas, el concepto de transitividad *parcial* puede resultar útil. En muchas redes, especialmente en redes sociales, el hecho de que u y v sean vecinos y de que v y w también lo sean no garantiza que exista una conexión entre u y v, pero sí lo hace mucho más probable. El amigo de mi amigo no necesariamente es mi amigo, pero es mucho más probable que él sea mi amigo, antes que un miembro de la población escogido al azar.

Una forma de cuantificar la transitividad en una red es la siguiente. Si u conoce a v y v conoce a w, decimos que estos tres nodos forman una *tríada*.

Si, además, u conoce a w, forman una *tríada cerrada*, o triángulo. Con estas definiciones en mente definimos el *coeficiente de clustering* C como la fracción de caminos de longitud dos en la red que son cerrados. En otras palabras,

$$C = \frac{3 \times \text{cantidad de tríadas cerradas}}{\text{cantidad de tríadas}}. \tag{2.4}$$

Es evidente que C está acotado entre 0 y 1, siendo $C = 1$ el caso de transitividad perfecta y $C = 0$ el caso donde no existen triángulos, como sucede en diversas topologías, como por ejemplo en un árbol o en una red regular cuadrada.

Como introdujimos anteriormente, es esperable que las redes sociales sean altamente transitivas. Esto puede verificarse computando el coeficiente de clustering para algunas redes. Por ejemplo, si consideramos la red de colaboraciones entre actores estudiada en [New03b], donde dos actores están conectados entre sí si participaron en una misma película, observamos que $C = 0{,}20$. Ahora bien, ¿cómo podemos saber si este valor es alto o bajo? Para esto, consideremos por ejemplo qué sucedería en una red del mismo tamaño en donde las conexiones entre nodos se dieran de manera aleatoria. Para hacerlo más simple, consideremos que en esta red todos los nodos tienen aproximadamente la misma cantidad c de conexiones. En ese caso, la probabilidad de que una dada tríada uvw esté cerrada es simplemente la probabilidad de que u y w estén conectados, la cual equivale a $c/(N-1)$, siendo N el tamaño de la red. Para el ejemplo de la red de actores, esta probabilidad es aproximadamente 0,0003, es decir, tres órdenes de magnitud menor. Si bien hay algunas suposiciones en juego, la diferencia es muy notoria y sirve además para ilustrar un procedimiento que repetiremos en distintos tipos de análisis, y que formalizaremos en la Sección 2.2. El mismo consiste en cuantificar la significancia estadística de de una métrica comparando el resultado obtenido en la red en estudio con el resultado obtenido en una red aleatoria de similares características.

El coeficiente de clustering introducido es una propiedad global de la red que cuantifica hasta qué punto los pares de nodos con un vecino en común se conectan entre sí. Sin embargo, a veces resulta útil pensar en una propiedad similar que sea local, o sea, que pueda calcularse para cada nodo individualmente. Con esta idea definimos el *coeficiente de clustering local* como

$$C_i = \frac{\text{cantidad de pares de vecinos de } i \text{ que están conectados}}{\text{cantidad de pares de vecinos de } i}. \tag{2.5}$$

Es decir, para calcular C_i es necesario iterar sobre todos los pares de vecinos de i, contar la cantidad de pares que están conectados entre sí y dividir ese número por el total de pares, el cual es $k_i(k_i-1)/2$, donde k_i es el grado de i. Para nodos con menos de dos vecinos la expresión (2.5) no está definida, por lo que usamos la convención $C_i = 0$.

A partir del coeficiente de clustering local, podemos definir un coeficiente

de clustering global como

$$C_{WS} = \frac{1}{N}\sum_{i=1}^{N} C_i. \tag{2.6}$$

Este coeficiente fue introducido por Watts y Strogatz en su célebre artículo de 1998 [WS98], lo que explica el subíndice escogido para la definición. Esta métrica es diferente a (2.4), pero ambas son ampliamente utilizadas en la literatura. En algunas familias de redes existe una alta correlación entre ambas, pero existen también casos donde sus valores son notoriamente distintos. Además, es frecuente observar en redes empíricas que $C_{WS} > C$. Esta desigualdad se debe a que el coeficiente de Watts-Strogatz tiende a estar dominado por nodos de bajo grado, abundantes en redes con distribución de grado heterogénea, los cuales contribuyen con altos valores de C_i al promedio (2.6).

Finalmente, como se indica en [VPV02], es a veces importante discriminar la caracterización de las redes de acuerdo al grado de los nodos. En este sentido, la dependencia del coeficiente de *clustering* con el grado de los nodos puede definirse como:

$$c(k) = \frac{1}{N_k}\sum_{i/k_i=k}^{N_k} C_{WS}(i), \tag{2.7}$$

donde la suma se extiende sobre todos los nodos en la red que tienen grado k y N_k es el número de tales nodos.

2.1.3. Homofilia y asortatividad

Asortatividad por atributos no ordinales

Supongamos que tenemos una red en la cual los nodos pueden ser clasificados por una característica que puede tomar una cantidad finita de valores, de carácter descriptivos y que no siguen un orden particular. Por ejemplo, podríamos considerar una red social en la que los miembros estén caracterizados por etnia, género o nacionalidad. Decimos que una red es *asortativa* cuando existe una fracción significativa de enlaces entre miembros que son del mismo tipo. Una manera de cuantificar esta idea es la siguiente. Para cada clase de nodos calculamos la fracción de enlaces que unen nodos de esa dada clase, y restamos a ese valor la cantidad de enlaces que habría en una red en la cual los enlaces fuesen distribuidos aleatoriamente. En términos matemáticos, podemos expresar esta idea de la siguiente manera. Supongamos que existen N_c clases de nodos, y que las numeramos como $1,\ 2,\ \ldots, N_c$ y sea g_i el grupo o clase a la cual pertenece el nodo i. La cantidad total de enlaces que unen nodos de una misma comunidad es

$$\sum_{\text{enlaces}(i,j)} \delta_{g_i,g_j} A_{ij} \delta_{g_i,g_j}, \tag{2.8}$$

donde δ_{ij} es la delta de Kronecker.

En términos de la matriz de adyacencia A del grafo (cuyos elementos A_{ij} valen 1 si el nodo i está conectado con j, y 0 en caso contrario), podemos reescribir la suma 2.9 de la forma

$$\frac{1}{2}\sum_{ij} A_{ij}\delta_{g_i,g_j}, \tag{2.9}$$

donde el factor $1/2$ viene del hecho de que la suma de la derecha cuenta dos veces cada par i, j.

Calculemos ahora la cantidad esperada de enlaces entre nodos para el caso en que los enlaces estén distribuidos de manera aleatoria, manteniendo la distribución de grado de la red. Sea i un nodo de grado k_i y consideremos un enlace e que tiene a i como uno de sus extremos. En una red con M enlaces existen $2M$ extremos de enlaces, por lo que la probabilidad de que el otro extremo de e sea uno de los k_j extremos de enlace salientes de un dado nodo j es $k_j/(2m-1) \simeq k_j/2m$. Contando ahora todos los k_i enlaces salientes de i, el valor esperado para la cantidad de enlaces entre los nodos i y j es

$$\frac{1}{2}\sum_{ij} \frac{k_i k_j}{2m}\delta_{g_i,g_j}. \tag{2.10}$$

La resta entre (2.10) y (2.9) nos da la diferencia entre la cantidad observada y la cantidad esperada de enlaces que unen nodos del mismo tipo:

$$\frac{1}{2}\sum_{ij} A_{ij}\delta_{g_i,g_j} - \frac{1}{2}\sum_{ij} \frac{k_i k_j}{2m}\delta_{g_i,g_j} = \frac{1}{2}\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{2m}\right)\delta_{g_i,g_j}. \tag{2.11}$$

Usualmente, en lugar de calcular la cantidad de enlaces, es conveniente calcular la fracción de enlaces. La magnitud resultante se denomina *modularidad* y puede expresarse como

$$Q = \frac{1}{2m}\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{2m}\right)\delta_{g_i,g_j}. \tag{2.12}$$

La modularidad fue introducida por Newman y Girvan en [NG04] en el contexto de detección de comunidades, tema que abordaremos en la Sección 2.3. Su valor está acotado por 1 y es cercano a cero cuando la red no exhibe segregación por tipo. Si bien en redes sociales suele ser mayor que cero, puede también tomar valores negativos, en caso de que haya menos enlaces dentro de cada grupo que lo esperado.

Asortatividad por atributos ordinales

Existen casos en que los atributos de los nodos tienen asociado un orden. Siguiendo con el ejemplo de las redes sociales, podemos tomar la edad o el

nivel de ingresos como dos atributos de este tipo. Al igual que en los ejemplos de atributos no ordinales discutidos en la sección anterior, las personas suelen relacionarse más con personas dentro de su mismo rango de edad (por ejemplo, los niños dentro de una escuela) o su nivel de ingresos (colegas de trabajo). Para este tipo de variables ordinales no resulta demasiado útil preguntarnos si las relaciones existen entre personas que tienen exactamente los mismos atributos, sino que basta con observar en qué medida los vínculos se dan con mayor frecuencia entre miembros con atributos similares. Frecuentemente, las redes sociales exhiben algún nivel de estratificación, como se representa en el esquema de la Figura 2.1. Una manera de cuantificar la asortatividad en este caso es mediante un análisis de covarianza. Sea x_i el valor del atributo X (de caracter escalar y ordinal) para el nodo i. Consideremos el par de valores x_i y x_j, correspondientes a los nodos i y j, ambos extremos de un dado enlace e y calculemos su covarianza respecto a todos los enlaces. Para ello, definimos primero el valor medio μ del valor x_i en el extremo de un enlace como

$$\mu = \frac{\sum_{ij} A_{ij} x_i}{\sum_{ij} A_{ij}} = \frac{1}{2m} \sum_i k_i x_i. \tag{2.13}$$

Luego, la covarianza entre x_i y x_j sobre enlaces es

$$\begin{aligned}
\mathrm{cov}(x_i, x_j) &= \frac{\sum_{ij} A_{ij}(x_i - \mu)(x_j - \mu)}{\sum_{ij} A_{ij}} \\
&= \frac{1}{2m} \sum_{ij} A_{ij}(x_i x_j - \mu x_i - \mu x_j + \mu^2) \\
&= \frac{1}{2m} \sum_{ij} A_{ij} x_i x_j - \mu^2 \\
&= \frac{1}{2m} \sum_{ij} A_{ij} x_i x_j - \frac{1}{(2m)^2} \sum_{ij} k_i k_j x_i x_j \\
&= \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) x_i x_j,
\end{aligned} \tag{2.14}$$

donde utilizamos la ecuación (2.13). Notemos la similitud entre esta expresión y la definición de modularidad (2.12). La única diferencia es el reemplazo de δ_{g_i, g_j} por $x_i x_j$.

Si normalizamos la expresión anterior para que la asortatividad perfecta corresponda al valor 1, obtenemos el denominado *coeficiente de asortatividad*

$$r = \frac{\sum_{ij} \left(A_{ij} - k_i k_j / 2m \right) x_i x_j}{\sum_{ij} \left(k_i \delta_{ij} - k_i k_j / 2m \right) x_i x_j}, \tag{2.15}$$

el cual corresponde al coeficiente de correlación de Pearson, que constituye una métrica estándar en la medición de correlación entre variables escalares. Este

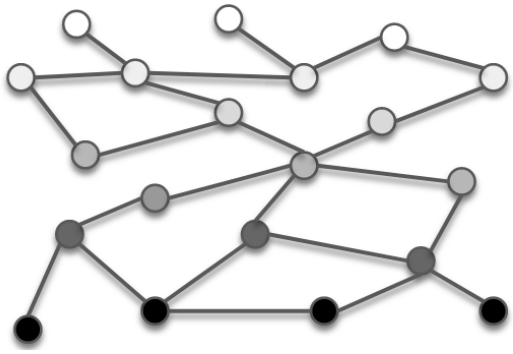

Figura 2.1: Esquema de estratificación en una red. Los nodos están caracterizados por un atributo ordinal representado por el sombreado de los nodos. Cada uno de los nodos de la red se vincula preferentemente con nodos de su mismo "nivel", o sea, aquellos que tienen valores similares de ese atributo.

coeficiente toma el valor 1 para una red perfectamente asortativa, y -1 para una red perfectamente disortativa. Un valor cercano a cero indica ausencia de correlaciones lineales, aunque podrían existir correlaciones no lineales no detectables por esta métrica.

Asortatividad por grado

Un caso de asortatividad por una variable ordinal que resulta de interés es la asortatividad por grado. En una red asortativa por grado, los nodos de grado alto se conectan preferentemente con otros nodos de grado alto, y lo mismo sucede con los de grado bajo.

La razón por la cual la asortatividad por grado es relevante es que, a diferencia de los casos mencionados anteriormente, el grado constituye una propiedad intrínseca de la red. El hecho de que una propiedad (en este caso el grado) determine otra (la ubicación de los enlaces) da lugar a aspectos interesantes en la red. Por ejemplo, en una red asortativa, los nodos de grado alto tienden a aglomerarse formando un núcleo (frecuentemente llamado *core*), rodeado de una periferia de nodos de baja conectividad. Esta estructura de núcleo-periferia se observa con frecuencia sobre todo en redes sociales, las cuales tienden a ser asortativas por grado [New02].

La asortatividad por grado puede determinarse utilizando el resultado (2.15), en donde la variable x_i es particularizada por el grado k_i. En ese caso, obtenemos la expresión

$$r = \frac{\sum_{ij} \left(A_{ij} - k_i k_j / 2m\right) k_i k_j}{\sum_{ij} \left(k_i \delta_{ij} - k_i k_j / 2m\right) k_i k_j}. \tag{2.16}$$

Un aspecto a tener en cuenta es que para evaluar (2.16) sólo necesitamos conocer la estructura de la red, sin ninguna información adicional. Una vez que tenemos la matriz de adyacencia de la red, conocemos los grados de todos los

nodos y por lo tanto podemos calcular r. Es por esta razón que la asortatividad por grado es el tipo de mezcla asortativa más frecuentemente estudiada en la literatura de redes.

Una manera alternativa de medir las correlaciones de grado es considerar la distribución de grado asociada a los vecinos de un nodo. Consideremos, por ejemplo, un nodo i con grado k_i, y definamos Γ_i al conjunto de vecinos de i. El grado medio de los vecinos de i es entonces

$$k_{\mathrm{nn},i} = \frac{1}{k_i} \sum_{j \in \Gamma_i} k_j. \tag{2.17}$$

Como se discute en [PVV01; VPV02], resulta conveniente promediar esta cantidad sobre los nodos que tienen el mismo grado. La cantidad resultante es

$$k_{\mathrm{nn}}(k) = \frac{1}{N_k} \sum_{i/k_i=k} k_{\mathrm{nn},i}, \tag{2.18}$$

donde N_k es la cantidad de nodos de grado k. Como se muestra en la Figura 2.2, una red asortativa está representada por un comportamiento creciente de k_{nn} respecto de k, mientras que lo opuesto ocurre para redes de naturaleza disortativa.

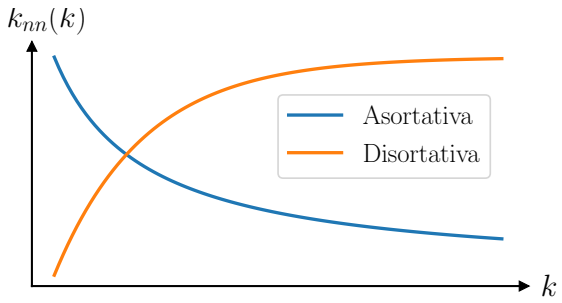

Figura 2.2: Representación esquemática de la propiedad de asortatividad de una red indicada por el comportamiento del grado medio de los primeros vecinos $k_{nn}(k)$ en función de k.

2.1.4. Centralidad

Uno de los aspectos que nos permiten ganar conocimiento sobre la estructura y dinámica de las redes complejas es la importancia o centralidad de sus componentes [Fre77]. La naturaleza heterogénea de las redes empíricas hace evidente el hecho de que los nodos y enlaces que la conforman tienen a asumir diferentes roles y consecuentemente pueden ser considerados más o menos importantes.

Identificar los nodos centrales no es una tarea simple. En particular, no existe un único criterio para determinar cuándo un nodo es más importante que otro.

¿Es más central un nodo con muchas conexiones o un nodo que articule diferentes grupos? La respuesta depende del sistema que estemos considerando, por lo que resulta imposible idear un índice universal que cuantifique la centralidad. Por otra parte, no sólo es necesario contar con una métrica apropiada para medir la importancia de los nodos, sino que además esa métrica tiene que ser computable en un tiempo acotado, algo que no es trivial en redes de gran tamaño.

Centralidad de grado

Probablemente la magnitud más simple con la que podemos medir la centralidad de un nodo sea su grado, es decir, la cantidad de conexiones que posee. A pesar de su simpleza, la centralidad de grado resulta en muchos casos una métrica adecuada para comparar nodos. En el contexto de redes sociales, por ejemplo, una persona con más contactos posiblemente tenga mayor influencia, prestigio o acceso a la información. En el ámbito científico, la cantidad de citas de un artículo se utiliza para determinar el impacto de la investigación científica que describe.

En términos operativos, esta centralidad resulta muy fácil de calcular. La complejidad computacional asociada a su cómputo es $\mathcal{O}(\langle k \rangle N)$, donde N es el tamaño de la red y $\langle k \rangle$ su grado medio. Esta baja complejidad es consecuencia del carácter local de esta medida. La centralidad de cada nodo depende exclusivamente de los vecinos más cercanos a ese nodo. Este aspecto hace también que esta medida sea robusta frente a cambios locales en la red. Por ejemplo, en el estudio de redes empíricas muchas veces no es posible conocer perfectamente la red, sino sólo una aproximación de la misma. Sin embargo, si esa aproximación es buena, entonces el cómputo de la centralidad de grado en la red aproximada también se aproximará al de la red verdadera. Como veremos más adelante, en las medidas de centralidad de carácter global esto no necesariamente sucede.

Centralidad de *betweenness*

Una medida de centralidad completamente diferente al grado es el *betweenness*, el cual mide qué tanto un nodo está ubicado sobre caminos que unen otros nodos[1]. La idea del *betweenness* es frecuentemente atribuida a Freeman en 1977 [Fre77] aunque, como el mismo Freeman menciona en [Fre04], fue propuesta algunos años antes por Anthonisse en un reporte técnico no publicado [Ant71]. Aunque comenzó en el contexto de redes sociales, su uso se ha esparcido por la literatura de redes complejas, con aplicaciones tales como la detección de comunidades [NG04], la robustez de las redes [Hol+02] y la organización de las ciudades [Kir+18].

[1] No parece haber una traducción literal de la palabra *betweenness* al español. Algunos libros de texto usan el término "intermediación", pero nosotros hemos escogido utilizar la nomenclatura en inglés.

Podemos definir la centralidad de betweenness de la siguiente manera. Sea $\sigma(s,t)$ la cantidad de caminos geodésicos (caminos de longitud mínima) que conectan los nodos s y t y sea $\sigma_i(s,t)$ la cantidad de dichos caminos que pasan por un dado nodo i. Definimos el betweenness del nodo i como

$$b_i = \sum_{s \neq t} \frac{\sigma_i(s,t)}{\sigma(s,t)}, \tag{2.19}$$

donde adoptamos la convención de que $\sigma(s,t)/\sigma(s,t) = 0$ cuando $\sigma(s,t)$ y $\sigma_i(s,t)$ son nulos.

El betweenness puede interpretarse como la cantidad de "carga" que pasa a través de un nodo cuando existe algún tipo de flujo sobre la red. Los nodos con betweenness alto articulan diferentes grupos de nodos y su importancia está directamente relacionada con la comunicación dentro del sistema.

La principal dificultad de trabajar con la centralidad de betweenness es el costo computacional requerido para su cómputo. Un algoritmo naïve requiere, para cada nodo, realizar una suma sobre todos los pares de nodos de la red. De esta forma, si N es la cantidad de nodos, la complejidad algorítmica es $\mathcal{O}(N^3)$. Afortunadamente, existe un algoritmo más eficiente, el cual fue desarrollado por U. Brandes en [Bra01]. Este algoritmo tiene complejidad $\mathcal{O}(NM)$, siendo M la cantidad de enlaces. Para el caso de redes esparsas, donde $M \sim N$, la mejora en eficiencia es significativa. Sin embargo, su complejidad sigue siendo cuadrática, lo cual dificulta su cómputo en grandes grafos.

Dada su gran utilidad y versatilidad, se ha dedicado mucho esfuerzo en poder calcular la centralidad de betweenness en grafos grandes. Una opción para reducir el tiempo de cómputo es paralelizar el algoritmo y emplear placas gráficas para sacar ventaja de esta paralelización. La otra alternativa es diseñar métodos de aproximación que permitan estimar el valor de betweenness empleando algoritmos menos costosos.

Variaciones del betweenness

La definición dada del betweenness no necesariamente representa la centralidad de los nodos de la mejor manera. Por ejemplo, el hecho de que sólo los caminos geodésicos sean tenidos en cuenta es de alguna manera restrictivo. Dependiendo del sistema, la información podría fluir no sólo por geodésicas sino también por caminos más largos. Por otra parte, es razonable pensar que para ciertas situaciones, como por ejemplo en el flujo de información en una red social, sólo sean relevantes los caminos cortos, por lo que puede resultar conveniente excluir los caminos de gran longitud del cálculo del betweenness. En esta dirección, existe una variante denominada ℓ-betweenness [BE06] (la cual será abordada en detalle en el Capítulo 5) que consiste en incluir en la suma de (2.19) sólo aquellos caminos cuya longitud no supere un dado valor ℓ. Es decir, para cada valor $\ell = 2,\ 3,\ \ldots$, tendríamos asociada la medida de

centralidad

$$b_i^{(\ell)} = \sum_{\substack{s \neq t \\ d(s,t) \leq \ell}} \frac{\sigma_i(s,t)}{\sigma(s,t)}, \tag{2.20}$$

es claro que si $\ell \geq D$, siendo D el diámetro de la red, entonces la versión acotada del betweenness equivale a la definición original. En el otro extremo, y para redes esparsas, se puede verificar que $b_i^{(2)} \sim k_i^2$.

En [ET10] los autores mostraron que las distribuciones de ℓ-betweenness para distintos ℓ obedecen una relación de escala característica que permite estimar el valor correspondiente de betweenness de largo alcance. Este comportamiento fue observado para una amplia variedad de sistemas, como por ejemplo redes aleatorias de tipo Erdös-Rényi, redes libres de escala, redes espaciales y redes empíricas. Estos resultados son interesantes desde el punto de vista teórico. El hecho de que haya comportamientos universales en las distribuciones de betweenness puede brindar información acerca de los mecanismos por los cuales se distribuye la carga en los sistemas complejos. Por otra parte, la relación de escala puede ser empleada para predecir tanto los valores de centralidad como el orden de centralidad de los nodos en redes grandes, donde no es posible calcular el valor exacto del betweenness debido a la complejidad computacional que dicho cálculo requiere [PC12]. La desventaja de este método, al igual que otros métodos de aproximación de betweenness, es que no es posible determinar, a priori, la magnitud del error cometido en cada aproximación.

2.1.5. El fenómeno "club de ricos"

El fenómeno "club de ricos" se refiere a la tendencia de los nodos de alto grado (los nodos *ricos*) a vincularse entre sí formando subgrafos densamente conectados (*clubs*) más fácilmente que los nodos de bajo grado. Los autores Zhou y Modragón [ZM04] introdujeron una medida para cuantificar esta tendencia, llamada *coeficiente de club de ricos*, y definida como

$$\phi(k) = \frac{2M_{>k}}{N_{>k}(N_{>k} - 1)}, \tag{2.21}$$

donde $M_{>k}$ es la cantidad de conexiones entre los $N_{>k}$ nodos remanentes después de remover los nodos (y sus conexiones) con grado menor o igual a un dado valor k, y $N_{>k}(N_{>k} - 1)/2$ es la máxima cantidad de conexiones que podrían existir entre los $N_{>k}$ nodos. En otras palabras, $\phi(k)$ es un coeficiente acotado entre 0 y 1 que representa la densidad del subgrafo inducido por los nodos de grado mayor a k. En el contexto de sistemas sociales, un coeficiente de club de ricos que incrementa con el grado corresponde a la existencia de una cierta forma de oligarquía en la organización del sistema.

Como mencionan Colizza et al. [Col+06], este coeficiente es una función monótonamente creciente aún en redes no correlacionadas, de manera que para verificar la existencia de un verdadero club de ricos es necesario compararla con

una obtenida a partir de un modelo nulo. la corrección de este efecto espurio se obtiene definiendo

$$\rho(k) = \phi(k)/\phi_{ran}(k), \tag{2.22}$$

donde ϕ_{ran} es el coeficiente de club de ricos de la red completamente aleatorizada con la misma distribución de grado que la original (veremos en la siguiente sección cómo construir este tipo de redes aleatorias). En este caso, un club de ricos verdadero corresponde a un cociente $\rho(k) > 1$.

2.2. Redes aleatorias y modelos nulos

Una red o *grafo aleatorio* es un modelo de red en el cual los valores de algunas propiedades están fijos, pero que en lo demás es aleatoria. En general, los modelos de grafos aleatorios no se definen en términos de redes individuales sino que se construyen *ensambles* de redes, es decir, distribuciones de probabilidad sobre redes posibles.

El estudio de grafos aleatorios tiene distintas aplicaciones. En primer lugar, el análisis estadístico de estos modelos puede brindar información útil a la hora de analizar la estructura y la dinámica de las redes complejas reales. Muchas veces resulta complicado trabajar con datos empíricos, ya que nos encontramos con diversas limitaciones (existencia y veracidad de los datos, completitud de las redes, etc.). Sin embargo, al trabajar con modelos podemos, en principio, disponer de una cantidad arbitraria de datos que nos permiten tener una buena estadística de las propiedades que estamos interesados en medir. Si existe una similitud entre los modelos empleados y las redes empíricas estudiadas, entonces podemos hacer algún tipo de inferencia que nos permita sacar conclusiones sobre el comportamiento de los sistemas reales. En segundo lugar, las redes aleatorias nos sirven como modelo nulo con el cual comparar los resultados de mediciones realizadas sobre redes empíricas. Si queremos determinar si una cantidad medida es significativa, resulta una buena práctica compararla con la misma medición realizada sobre un ensamble de redes aleatorias "similar" a la red estudiada, donde el concepto de similaridad depende del sistema particular que estemos estudiando.

2.2.1. Modelo de Erdös-Rényi

El modelo más ampliamente estudiado de redes aleatorias se conoce como modelo de Erdös-Rényi, ya que fue popularizado por los matemáticos Paul Erdös y Alfréd Rényi en una serie de artículos publicados entre 1950 y 1960 [ER59; ER60]. Este modelo de redes se define mediante el ensamble $G(N, p)$, constituido por todas las redes de tamaño N en las cuales cada enlace está presente con una dada probabilidad p. Como presencia de cada enlace es completamente independiente de la existencia de otros enlaces, este tipo de redes carece de cualquier tipo de correlaciones entre nodos.

La simpleza de este modelo de redes hace posible el cálculo de muchas cantidades de interés. Por ejemplo, la cantidad promedio de enlaces en la red es

$$\langle m \rangle = \binom{N}{2} p. \tag{2.23}$$

De esta expresión podemos obtener también el grado medio $\langle k \rangle$ como

$$\langle k \rangle = \frac{2\langle m \rangle}{N} = \frac{2}{N}\binom{N}{2} p = (N-1)p \simeq Np. \tag{2.24}$$

Cuando analizamos redes empíricas, típicamente tratamos con sistemas en los cuales el grado medio tiene un valor acotado, independientemente del tamaño de la red. Para que esto se cumpla, es necesario considerar una probabilidad de conexión p dependiente de N en la forma $p \sim N^{-1}$.

No sólo el grado medio puede calcularse explícitamente en este modelo, sino también la distribución de grado p_k. Para determinarla basta notar que, dado que las conexiones son completamente aleatorias, la probabilidad de que un nodo se conecte con k nodos en particular es $p^k(1-p)^{N-1-k}$. Como hay $\binom{N-1}{k}$ formas de elegir estos nodos, la probabilidad de que el nodo en cuestión tenga exactamente k vecinos es

$$p_k = \binom{n-1}{p} p^k (1-p)^{N-1-k}. \tag{2.25}$$

En otras palabras, la distribución de grado del modelo $G(N,p)$ es binomial. En el caso en el que N sea grande y $\langle k \rangle$ se mantenga acotado, puede demostrarse fácilmente que la expresión (2.25) puede ser aproximada por una distribución de Poisson de la forma

$$p_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}, \tag{2.26}$$

motivo por el cual los grafos de Erdös-Rényi reciben habitualmente el nombre de *grafos aleatorios de Poisson*.

2.2.2. Modelo de configuración

Aunque el modelo de Erdös-Rényi discutido en la sección anterior ha demostrado ser de enorme utilidad en el estudio de las redes complejas, posee algunas limitaciones. La principal desventaja es que la distribución de grado sigue una ley de Poisson, que difiere significativamente de las distribuciones sesgadas que exhiben muchas redes empíricas. En esta sección discutiremos el modelo de configuración [BC78; MR95], un modelo un poco más sofisticado que puede tener cualquier distribución de grado, pero manteniendo la ventaja de que muchas de sus propiedades pueden ser calculadas explícitamente en el límite de grandes redes.

Estrictamente, el modelo de configuración es un modelo de grafos aleatorios con una dada *secuencia de grado*, en lugar de una distribución de grado. Es

decir, el ensamble está compuesto por grafos en donde cada nodo tiene una cantidad exacta de vecinos. Como consecuencia, también la cantidad de enlaces está determinada de manera exacta y es igual a $m = \frac{1}{2}\sum_i k_i$, a diferencia del modelo de Erdös-Rényi, en donde m es una variable aleatoria.

Para comprender el modelo de configuración, supongamos que especificamos el grado k_i de cada nodo $i = 1, \cdots, N$ de nuestra red y observemos el proceso descripto en la Figura 2.3. A cada nodo i asignamos un total de k_i "extremos de enlaces". Luego, elegimos dos de estos extremos uniformemente al azar y los conectamos entre sí formando un enlace completo, como se indica en la línea punteada de la figura. Repetimos este paso, eligiendo de a pares de extremos, hasta que no haya más extremos disponibles. El resultado es una red en la cual cada nodo tiene la cantidad deseada de vecinos.

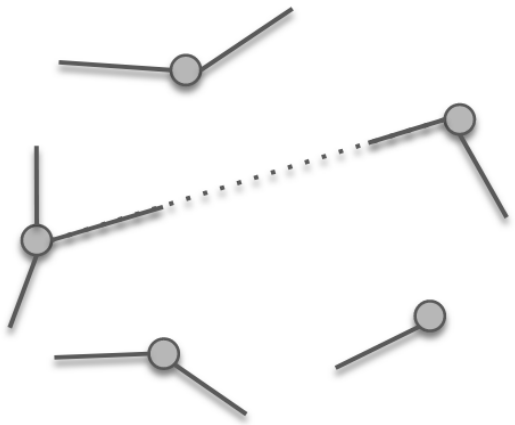

Figura 2.3: Esquema del modelo de configuración. Para cada nodo asignamos un conjunto de extremos de enlaces igual a su grado. Luego, los pares de extremos son elegidos de manera aleatoria y conectados formando enlaces completos.

Dado que no existen restricciones a la hora de elegir los extremos de los enlaces, el resultado final del algoritmo puede ser un grafo con enlaces múltiples o con auto-conexiones. Una forma de evitar esto sería descartar cada elección de pares de extremos que diera lugar a este tipo de conexiones. Sin embargo, esta variante del algoritmo ya no generaría de forma uniforme los grafos, por lo que en la práctica resulta una mala idea. Por otra parte, el modelo de configuración no nos asegura que el grafo generado sea conexo, propiedad que es deseable en determinadas condiciones. Afortunadamente, existe una variante de este modelo, introducida por Viger y Latapy en [VL15], con la cual es posible generar, de manera aleatoria y con probabilidad uniforme, grafos simples y conexos con una secuencia de grado específica.

2.2.3. Intercambio de conexiones

Existe, además del modelo de configuración, una estrategia alternativa para generar redes aleatorias con una secuencia de grado específica, conocida como

recableado o *intercambio de conexiones*. Esta estrategia consiste en lo siguiente. Supongamos que partimos de una red empírica, y queremos construir una red aleatoria que mantenga su misma secuencia de grado. Elegimos al azar dos enlaces disjuntos (sin nodos en común), $e_1 = (v_1, w_1)$ y $e_2 = (v_2, w_2)$. Luego, si no existen conexiones entre los extremos de e_1 y los de e_2, reemplazamos los enlaces por $e'_1 = (v_1, v_2)$ y $e'_2 = (w_1, w_2)$. Es fácil comprobar que esta transformación no altera la secuencia de grado de la red. Si repetimos este procedimiento muchas veces, las correlaciones entre nodos de la red original irán disminuyendo hasta alcanzar la máxima aleatoriedad. Puede probarse que este proceso es ergódico [Ors+15], lo que asegura un buen muestreo del ensamble de redes con secuencia de grado fija, siempre que se ejecuten suficientes pasos. De hecho, de acuerdo con Milo, et al. [Mil+03], es suficiente con realizar una cantidad de intercambios igual a diez veces la cantidad de enlaces de la red para alcanzar la máxima aleatoriedad.

2.2.4. Triangulación de Delaunay

La organización de muchos sistemas complejos se desarrolla bajo restricciones espaciales. Las redes de distribución eléctrica, las sistemas de transporte y movilidad urbana, Internet y el cerebro humano son todos ejemplos de sistemas cuya estructura y evolución son influenciados por aspectos geométricos. Las redes espaciales, en las cuales los nodos y los enlaces están embebidos en espacios de dos o tres dimensiones, constituyen un modelo natural para estudiar estos sistemas [Bar11].

El modelado realista de redes embebidas se realiza frecuentemente empleando diferentes modelos de redes espaciales aleatorias. En la literatura de la física, uno de los modelos más frecuentemente estudiados es la teselación de Voronoi y su dual, la triangulación de Delaunay (DT), también llamada *random lattice*[2] [BZ09; Kir+18]. Vale la pena mencionar que la distribución de algunas de las medidas locales introducidas en la Sección 2.1, como el grado, la transitividad y la asortatividad, se vuelven menos informativas en el caso de redes espaciales. Por ejemplo, las redes planares frecuentemente exhiben una distribución de grado centrada, mientras que las triangulaciones maximizan trivialmente la transitividad. En contraste, otras medidas tales como las distancias topológicas y geométricas se vuelven más relevantes para este tipo de redes.

La triangulación de Delaunay puede definirse de la siguiente manera. Dado un conjunto de puntos V en un espacio de dimensión d (los nodos de la red), se establece un enlace entre dos puntos $u, v \in V$ sí, y sólo sí, ambos pueden ser inscriptos en una esfera $d-$dimensional que no incluya a ningún otro punto del conjunto, como se ilustra en la Figura 2.4 La DT es un ejemplo de *grafo de región excluida* [KBD19], donde la conectividad entre dos nodos se establece

[2]La palabra *lattice* suele traducirse al español como "red regular". Sin embargo, no sería adecuado llamar "red regular aleatoria" a las redes de Delaunay ya que las mismas no son grafos regulares.

en base a la ausencia de puntos en una dada región entre dichos nodos. Otros miembros de esta clase de redes son los grafos de Gabriel, los grafos de vecinos relativos y los árboles de expansión mínima euclídeos, todos subgrafos de la DT [KBD19].

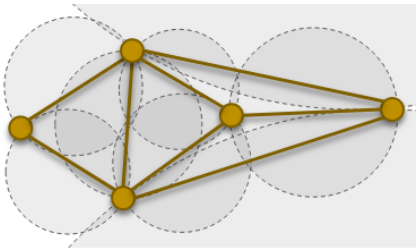

Figura 2.4: Triangulación de Delaunay de un conjunto de 5 puntos ubicados sobre el plano.

El estudio de las redes de Delaunay tiene numerosas aplicaciones, que van desde el diseño de redes inhalámbricas ad-hoc [Meg+01], hasta el modelado de ciudades [Kir+18]. Desde el punto de vista de la mecánica estadística, la DT puede considerarse como una versión perturbada de las redes regulares con la presencia de desorden topológico o "congelado" [JW04; BV14], por lo que resulta natural emplearla como modelo para estudio de transiciones de fase en sistemas con desorden de este tipo [JV95; Lim+00; DAF16; McC87]. Como veremos en el Capítulo 5, la existencia de esta clase de desorden puede modificar de manera sustancial la dinámica de los procesos, generando transiciones de fase diferentes a las observadas en redes regulares.

2.3. Comunidades

En muchas redes complejas la distribución de conexiones tiende a ser inhomogénea en diferentes niveles de organización. En un nivel microscópico, las inhomogeneidades se manifiestan en las colas largas que aparecen en la distribución de grado. En una escala mesoscópica, sucede frecuentemente que ciertos grupos de nodos se encuentran densamente conectados entre ellos y relativamente menos conectados con el resto. Estos grupos de nodos se dice que forman comunidades, también llamadas *clusters* o módulos. Las comunidades pueden aportar información importante sobre la red, ya que es muy probable que nodos pertenecientes a una misma comunidad tengan propiedades en común o cumplan una función similar dentro la red [HPF16]. Dada una red y una partición en particular, es decir, una función que asigna a cada nodo una comunidad, uno puede calcular la modularidad Q [CNM04], definida por la expresión (2.12). Esta propiedad cuantifica en que medida una red se encuentra dividida en comunidades, evaluando las conexiones de los nodos que se encuentran en una misma comunidad (conexiones internas). Si la fracción de conexiones internas no difieren de la que se espera para una red aleatorizada, el

valor de la modularidad es cero. De acuerdo a Clauset *et al.* [CNM04] un valor de Q mayor a 0,3 indica que la red tiene una estructura modular significativa. Otra cantidad importante, que permite evaluar si las comunidades en una red se encuentran bien definidas, es el parámetro de mezcla μ [For10], definido como

$$\mu = \frac{\sum_i k_i^{ext}}{2M}, \tag{2.27}$$

donde M es la cantidad de enlaces y k_i^{ext} es el grado externo del nodo i, es decir, la cantidad de vecinos de i que pertenecen a comunidades diferentes a la suya. Dicho de otro modo, μ representa la fracción de enlaces que unen nodos de comunidades distintas, por lo que las comunidades estarán mejor definidas para valores bajos de este parámetro.

2.3.1. Maximización de la modularidad

Una de las estrategias empleadas para la detección de comunidades es definir un problema de optimización en donde se busque maximizar una función objetivo sobre todas las posibles particiones de la red. Una elección razonable de la función objetivo es la modularidad. Desafortunadamente, maximizar esta función es un problema que pertenece a la categoría *NP-hard* [Bra+08], por lo que sólo es posible encontrar soluciones aproximadas mediante algoritmos heurísticos.

El algoritmo Louvain [Blo+08], que lleva el nombre de la ciudad belga de residencia de sus autores, es uno de los algoritmos heurísticos que ha sido desarrollado para maximizar de manera aproximada la modularidad. Es uno de los más ampliamente empleados en la literatura de detección de comunidades, ya que se trata de un algoritmo fácil de computar y que ha demostrado dar buenos resultados en la práctica. El algoritmo funciona de la siguiente manera. Inicialmente, consideramos que cada nodo de la red constituye una comunidad en sí misma. A partir de esto, hacemos un barrido sobre los nodos. Para cada nodo i, calculamos la diferencia en modularidad que habría si tal nodo se uniera a la comunidad de su vecino j. Repetimos esto para cada uno de los vecinos de i y, si la mayor de esas diferencias es positiva, unimos a i con el vecino correspondiente. Al final del barrido obtenemos el primer nivel de la partición. Una vez hecho esto, construimos una red pesada en donde los nodos corresponden a las comunidades previamente encontradas, y los pesos de las conexiones equivalen a la cantidad de enlaces entre nodos de cada comunidad en la red original. Una vez obtenida esta nueva red repetimos el paso de barrido, agrupando nuevamente los nodos hasta obtener el siguiente nivel de partición. Dado que ahora nuestra red es pesada, en lugar de utilizar la matriz de adyacencia en la definición de modularidad, utilizamos la matriz de pesos W, donde el elemento W_{ij} indica el peso de la conexión entre el nodo i y el nodo j. Esta secuencia de dos pasos (maximización local de la modularidad y aglomeración de nodos) se repite hasta que no sea posible

seguir aumentando la modularidad. El resultado entonces es una jerarquía de particiones. Si sólo estamos interesados en hallar la partición que maximice Q, entonces nos quedamos con la última partición hallada por el algoritmo (la más gruesa). Sin embargo, vale la pena resaltar que la jerarquía puede ser de interés de por sí, ya que nos brinda una descripción de la estructura de comunidades de la red a distintos niveles de detalle.

Si bien el algoritmo de Louvain es eficiente y, por lo general, efectivo, recientemente se ha visto que tiene algunos problemas. En [TWE19], los autores muestran que existen situaciones en las que Louvain encuentra comunidades arbitrariamente mal conectadas. En el caso extremo, asigna una misma comunidad a conjuntos disconexos de nodos. Este inconveniente no sólo es posible en la teoría, sino que sucede frecuentemente en la práctica, y se agrava aún más si el número de iteraciones del algoritmo aumenta. Para remediar este inconveniente, los autores idearon una variante, un poco más compleja, del algoritmo, a la que le dieron el nombre de Leiden, por su ciudad de residencia. Como muestran en [TWE19], Leiden soluciona el problema de comunidades disconexas. Más aún, el algoritmo es más rápido que Louvain y encuentra, en la mayoría de los casos, valores ligeramente más altos para la modularidad.

2.3.2. Comunidades basadas en atributos de los nodos

Cuando introdujimos el concepto de asortatividad en la Sección 2.1.3, mencionamos que el mismo concepto puede ser empleado tanto para cuantificar correlaciones entre propiedades estructurales de las redes, tales como el grado, como para estudiar atributos de los nodos, tales como la edad o el género de las personas en una red social. Para el caso de detección de comunidades, hemos discutido hasta ahora algoritmos que se basan exclusivamente en variables topológicas de las redes. No obstante, es natural pensar que los atributos de los nodos pueden brindar información relevante a la hora de identificar la estructura modular de una red. Uno de los algoritmos que fueron desarrollados con esta idea fue introducido por Newman y Clauset en [NC16]. Este algoritmo hace uso de la inferencia bayesiana para construir un modelo generativo de redes. Las redes generadas poseen algunas de las características de las redes que se quiere emular, tales como la estructura de comunidades y el atributo de los nodos. Es así que el modelo se ajusta a las redes con atributos observada y los parámetros del ajuste dan información acerca de la estructura de la red. El modelo se basa en el *stochastic block model* [HLL83], donde se incorpora la dependencia de los nodos con los atributos usando un conjunto de probabilidades *anteriores*. En importante recalcar que este algoritmo no supone *a priori* ninguna relación entre estructura y atributos, pero si esta existe el algoritmo debería ser capaz de evidenciarla. Por otra parte, el algoritmo presenta dos limitaciones. En primer lugar, el número de comunidades que se pretende detectar es fijado de antemano con un parámetro, en contraste con otros algortitmos tales como el de Louvain. Segundo, el modelo desarrollado por Newman y Clauset está

limitado a atributos categóricos no ordinales. En caso de tratar con un atributo ordinal y continuo, es necesario transformarlo en un conjunto de categorías, que pueden definirse a través de rangos. El algoritmo luego interpretará cada una de esas categorías de manera independiente, perdiendo la relación de orden original.

Para determinar la correlación entre las categorías de atributo y las comunidades encontradas, los autores recomiendan utilizar la información mutua normalizada [TT91], la cual se define como

$$\mathrm{NMI} = \frac{I(\mathbf{s}, \mathbf{x})}{\min\left[H(\mathbf{s})H(\mathbf{x})\right]}, \tag{2.28}$$

donde $I(X, Y)$ es la información mutua entre las variables X e Y, y $H(X)$ denota la entropía. El vector $\mathbf{x} = [x_1, \ldots, x_N]$ indica la categoría de atributo correspondiente a cada nodo, mientras que $\mathbf{s} = [s_1, \ldots, s_N]$ indica la pertenencia de cada nodo a una de las comunidades determinadas por el algoritmo. La información mutua normalizada está acotada entre 0 y 1, y toma un valor mayor cuanto mayor sea la correlación entre las categorías del atributo y estructura de comunidades predicha por el algoritmo. Si bien esta métrica es ampliamente utilizada, existen otras opciones que pueden resultar más adecuadas, contemplando, por ejemplo, correcciones por efectos aleatorios [VEB09] o extensiones a particiones jerárquicas [PAS20].

2.4. Percolación y robustez de las redes

Hasta ahora, nuestro marco teórico se ha basado en diferentes formas de caracterizar la estructura de las redes complejas. Si bien esto es de gran importancia, constituye sólo un un paso hacia la comprensión de los sistemas complejos. En particular, resta responder cuál es la relación entre la estructura de una red y su funcionalidad. En otras palabras, supongamos que analizamos diferentes aspectos estructurales de una red. ¿Qué predicciones podemos hacer sobre el comportamiento global del sistema? Desafortunadamente, no existe una teoría unificada que nos permita responder a esta pregunta. De hecho, probablemente no exista una única respuesta posible, sino variadas respuestas, dependientes del sistema que estamos estudiando. Sin embargo, existen algunas áreas de investigación en donde se ha hecho un progreso notable en esta dirección. Una de ellas es el estudio del fenómeno de percolación, el cual da lugar a una teoría elegante sobre la robustez de las redes complejas frente a fallas en sus componentes.

2.4.1. Daños progresivos

La forma más simple de evaluar el comportamiento de las redes en el caso de daños progresivos es estudiar el efecto de la remoción de nodos. Aunque este análisis se enfoca sólo en el impacto del daño sobre la topología de la

red y omite todos los aspectos relacionados con los aspectos técnicos de los elementos del sistema y de su arquitectura, permite tener una intuición sobre los potenciales efectos de las fallas en redes de gran escala.

Existen dos escenarios cualitativamente distintos que podemos imaginar al considerar la remoción progresiva de nodos, ambos esquematizados en la Figura 2.5. El más simple de ellos es aquel en el cual los nodos son removidos de forma aleatoria. Este escenario permite modelar *fallas* en el sistema y constituye una buena aproximación para describir numerosas situaciones que pueden ocurrir en redes empíricas. En la red de Internet, por ejemplo, rara vez observamos todos los servidores en funcionamiento al mismo tiempo. Lo más frecuente es que una fracción (reducida) de los mismos esté sin servicio, aunque sin comprometer el funcionamiento global del sistema de manera significativa. El segundo escenario constituye los casos en los cuales los nodos son escogidos con la idea de emular algún tipo de daño intencional sobre la red. En estos *ataques dirigidos*, los nodos son típicamente removidos preferencialmente de acuerdo con alguna medida de centralidad, con la idea de generar el mayor daño posible sobre la red.

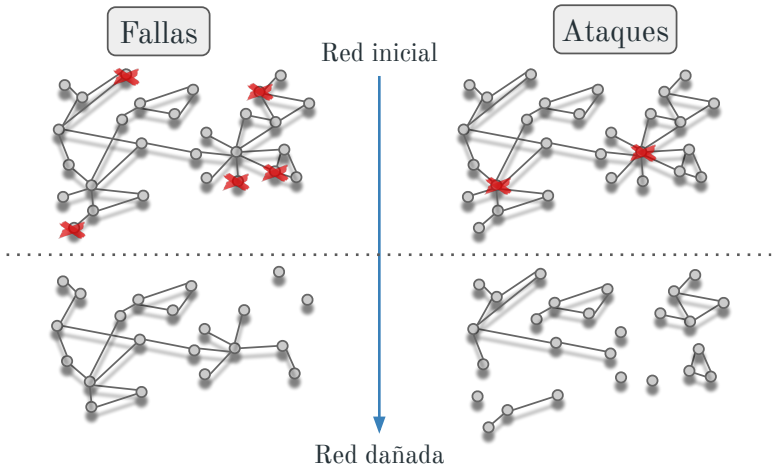

Figura 2.5: Comparación esquemática entre fallas aleatorias y ataques dirigidos. A la izquierda, los nodos son removidos de manera aleatoria, mientras que a la derecha se eliminan los nodos de mayor grado. Podemos apreciar que el daño producido por el ataque es considerablemente mayor el que el generado por las fallas.

2.4.2. Percolación

Para caracterizar la fenomenología asociada a los procesos de daños progresivos necesitamos definir una medida cuantitativa del daño. Para ello, consideremos una red inicialmente conexa en la cual una fracción f de los nodos ha sido removida (los enlaces correspondientes a los nodos removidos son eliminados también). Este daño podría romper la red, generando varias componentes conexas. La métrica cuantitativa más simple que podemos emplear para expresar el

daño es el tamaño relativo de la componente más grande de la red resultante, $S_1(f) = N_1(f)/N$, donde N es el tamaño de la red y $N_1(f)$ es el tamaño de la red luego de remover fN nodos. La red podrá mantener su funcionalidad mientras exista una componente de tamaño comparable con el de la red en su totalidad. Si $S_1 \ll 1$, significa que la red ha sido rota en muchas partes pequeñas, por lo que ya no es funcional.

La evolución de $S_1(f)$ depende del tipo y del tamaño de la red, así como también del tipo de daño que se esté considerando. Sin embargo, existe un patrón que se repite en numerosas situaciones y es el que ilustramos en la Figura 2.6. Si la red dañada es relativamente pequeña (panel (a)), $S_1(f)$ típicamente disminuye de manera progresiva, por lo que no existe una clara distinción entre el régimen de red "funcional" y "no funcional". Cuando el tamaño del sistema aumenta, el cambio de comportamiento se vuelve más preciso. Como podemos ver en el panel (b) de la figura, en el límite de tamaño infinito aparece una transición de fase continua a una dada fracción crítica de nodos removidos f_c. Cuando $f > f_c$, todas las componentes de la red tienen tamaño finito, por lo que $S_1 = 0$. En cambio, si $f < f_c$ la componente más grande escala linealmente con el tamaño del sistema, de manera tal que $S_1 > 0$. Por este comportamiento, esta componente recibe el nombre de *componente gigante*. Mientras existe una componente gigante, decimos que el sistema está en su estado *percolado*, y damos al punto crítico f_c el nombre de *umbral de percolación*.

El umbral de percolación es una de las características distintivas en una transición de fase de percolación. La determinación precisa de este observable es de gran importancia, aunque su interpretación está ligada al sistema que estemos estudiando. Dependiendo del caso, este umbral nos podría estar indicando la fracción mínima de personas que es necesario vacunar para mitigar un brote epidémico [MN00], la cantidad mínima de transformadores que deben estar en funcionamiento para garantizar un correcto servicio de distribución de energía eléctrica [Hol+02], o la máxima cantidad de neuronas que pueden fallar antes de que la funcionalidad de un sector del cerebro se vea comprometida [Mor+17].

La gráfica de la Figura 2.6b sugiere que, en un entorno del punto crítico y en el *límite termodinámico* ($N \to \infty$), la forma funcional de S_1 puede describirse de la forma

$$S_1 = \begin{cases} 0 & f > f_c, \\ a(f_c - f)^\beta & f \leq f_c, \end{cases} \tag{2.29}$$

donde a es una constante de proporcionalidad y $\beta > 0$ es el *exponente crítico* asociado a S_1. La transición entre el estado percolado y no percolado se ha estudiado muy extensamente en la física estadística y en muchos modelos de redes se observa que esta transición es continua. En este marco teórico S_1 se considera el *parámetro de orden* de la transición y f *recibe el nombre de parámetro de control*. Como es común en las transiciones continuas, otras cantidades también manifiestan un comportamiento crítico cerca del umbral de percolación. Una de estas medidas es el tamaño medio de las componentes, que

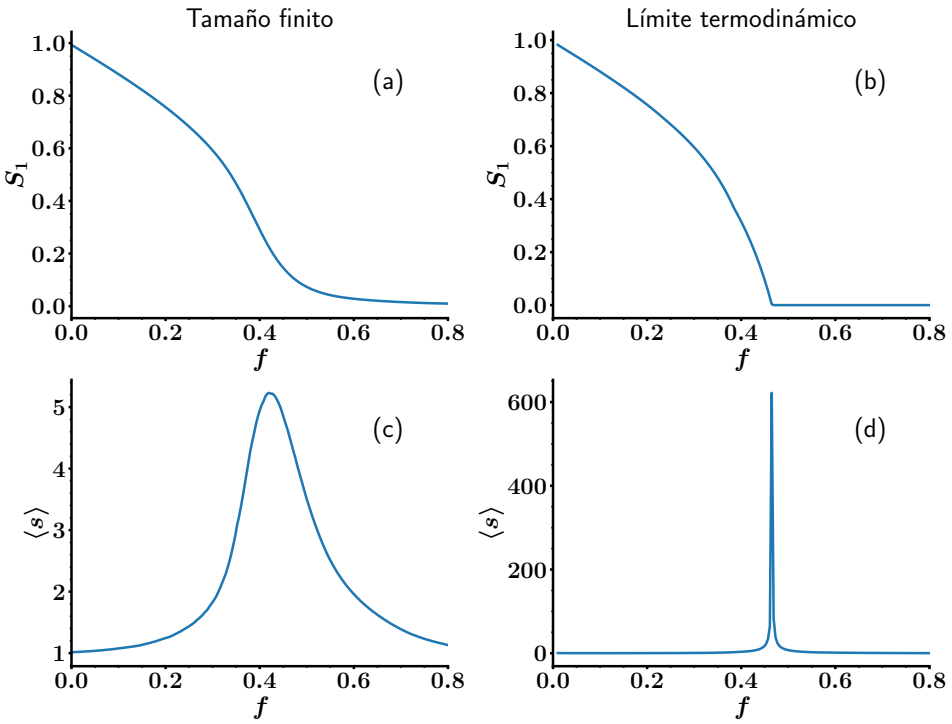

Figura 2.6: Gráfico esquemático para una transición de percolación continua sobre redes complejas. **(a-b)** Tamaño relativo de la componente gigante en un sistema de tamaño finito y en el límite termodinámico. **(c-d)** Tamaño medio de las componentes finitas en una red de tamaño finito y en el límite termodinámico.

juega el rol de *susceptibilidad* y se calcula como,

$$\langle s \rangle = \frac{\sum_s{}' s^2 n_s}{\sum_s{}' s n_s} \tag{2.30}$$

donde $n_s(f)$ es el número de componentes de tamaño s por nodo y la suma primada excluye la componente gigante. Como podemos apreciar en la Figura 2.6c, $\langle s \rangle$ presenta un máximo en torno a f_c. La altura de este máximo aumenta en tanto aumentamos el tamaño del sistema, y en el límite termodinámico diverge de la forma $\langle s \rangle \sim |f_c - f|^{-\gamma}$, donde $\gamma > 0$ es otro exponente crítico (panel (d)). Además, $n_s(f)$ tiene su propio comportamiento crítico cerca de f_c, en donde se vuelve heterogénea y puede ser descripta por la expresión,

$$n_s(f) \sim s^{-\tau} e^{-s/s^*}. \tag{2.31}$$

En esta expresión, s^* representa el tamaño de componente característico, que escala como $s^* \sim |f_c - f|^{-1/\sigma}$. Luego, en el punto de percolación, el número de de componentes de tamaño s sigue una ley de potencias $n_s(f) \sim s^{-\tau}$. Finalmente, la longitud de correlación ξ, la cual define la longitud de componente típica, diverge como $\xi \sim |f_c - f|^{-\nu}$, donde $\nu > 0$ [SA18].

La teoría de los fenómenos críticos establece que las transiciones de fase continuas (y, en ciertos casos, también las discontinuas) pueden ser caracterizadas completamente por sus exponentes críticos. Si dos sistemas comparten los mismos exponentes, se dice que pertenecen a la misma *clase de universalidad*. En percolación aleatoria, sólo dos exponentes son independientes. Los demás pueden obtenerse mediante las llamadas relaciones de escala. Por ejemplo, el exponente de Fisher τ puede obtenerse combinando los exponentes asociados al parámetro de orden y a la susceptibilidad mediante [SA18]

$$\tau = 2 + \frac{\beta}{\gamma + \beta}. \tag{2.32}$$

Como γ y β son ambos no negativos, entonces se deduce de (2.32) que $\tau \geq 2$. Otra relación que resulta de utilidad es [Bra96; FR11]

$$\frac{2\beta}{d\nu} + \frac{\gamma}{d\nu} = 1. \tag{2.33}$$

Esta última, conocida como relación de hiperescala, involucra la dimensión d del sistema. En el caso de redes complejas, que no están embebidas en un espacio d-dimensional, este parámetro representa la dimensión efectiva del sistema. Por ejemplo, la dimensión efectiva de las redes de Erdös-Rényi para el proceso de percolación aleatoria es $d = 6$.

2.4.3. Teoría de escaleo de tamaño finito

Como mencionamos en la sección anterior, los parámetros que definen la clase de universalidad de una transición continua son sus exponentes críticos. La determinación analítica de estos exponentes suele ser una tarea díficil; sólo se conocen con exactitud los exponentes correspondientes a algunos modelos en particular. En la mayoría de los casos, lo mejor que podemos hacer es aproximar el valor de los exponentes mediante simulaciones numéricas. Para ello, suele emplearse la *teoría de escaleo de tamaño finito* (FSSA), herramienta que ha demostrado ser muy efectiva en esta tarea [SA18; Sta87].

Para introducir la FSSA, consideremos un sistema de tamaño $N = L^d$, donde L es la dimensión lineal del sistema y d su dimensión[3]. Sea X una variable termodinámica del sistema. En un entorno cercano a la transición de percolación, podemos expresar la forma funcional de X como

$$X \sim g(t, L), \tag{2.34}$$

donde $t = (f - f_c)/f_c$ es el parámetro de control reducido. De acuerdo con la hipótesis de escala, la llamada *función de escala* g es continua y típicamente

[3]En redes complejas que no están embebidas en un espacio d-dimensional, el parámetro d corresponde a la dimensión efectiva, mientras que la longitud L no necesariamente tiene una interpretación física.

satisface una relación de homogeneidad de la forma [Sta87]

$$\lambda g(t, L) = g(\lambda^{-\omega} t, \lambda^{-\alpha} L), \tag{2.35}$$

para todo t y L. En particular, si elegimos $\lambda^{\alpha} = L$, entonces la ecuación (2.35) queda

$$L^{1/\alpha} g(t, L) = g(L^{-\omega/\alpha}, 1). \tag{2.36}$$

Definiendo la función de escala univaluada $G(x) \equiv g(x, 1)$, nos queda

$$g(t, L) = L^{-1/\alpha} G(L^{-\omega/\alpha} t). \tag{2.37}$$

Como la longitud de correlación diverge como $\xi \sim t^{-1/\nu}$ (ver Sección 2.4.2), podemos escribir

$$g(t, L) = L^{-1/\alpha} G(L^{-\omega/\alpha} \xi^{-\nu}). \tag{2.38}$$

Ahora bien, el argumento de la función de escala debe ser adimensional, lo cual impone la relación entre exponentes $\omega/\alpha + \nu = 0$. Utilizando este hecho y definiendo $F(x) \equiv G(x^{\nu})$, podemos expresar

$$\begin{aligned} g(t, L) &= L^{-\omega/\nu} G(L^{\nu} t) \\ X &\sim L^{-\omega/\nu} F(L t^{1/\nu}). \end{aligned} \tag{2.39}$$

Además, si tomamos el límite termodinámico al tiempo en que nos mantenemos cerca de la transición ($L \to \infty$ tal que $L t^{1/\nu} = 0$) y utilizamos la relación de escala para la longitud de correlación, obtenemos

$$X \sim L^{-\omega/\nu} \sim \xi^{-\omega/\nu} \sim t^{\omega}. \tag{2.40}$$

Por último, notemos que las expresiones anteriores pueden escribirse en términos de tamaño del sistema N, en lugar de hacerlo en términos del tamaño lineal L. En ese caso, la expresión (2.39) toma la forma

$$X \sim N^{-\omega/\bar{\nu}} F(N t^{1/\bar{\nu}}), \tag{2.41}$$

con $\bar{\nu} = d\nu$.

Un caso de uso de la ecuación (2.39) es el escaleo asociado al tamaño relativo de las componentes

$$S_i \sim L^{-\beta/\nu} \tilde{S}_i[(f - f_c) L^{1/\nu}]. \tag{2.42}$$

Aquí el subíndice $i = 1, 2, \ldots$ identifica cada componente, ordenadas por tamaño en forma descendiente. En particular, estaremos interesados en el parámetro de orden S_1 y en el tamaño absoluto de la segunda componente $S_2 L^d$. Otra relación de escala frecuentemente empleada es la asociada al tamaño medio de las componentes finitas,

$$\langle s \rangle (f, L) \sim L^{\gamma/\nu} \tilde{S}[(f - f_c) L^{1/\nu}]. \tag{2.43}$$

En un sistema de tamaño finito, el punto de percolación no necesariamente coincide con el correspondiente para $L \to \infty$. En general, la diferencia entre estos valores presenta un escaleo de la forma

$$f_c(L) - f_c = bL^{-\lambda}, \tag{2.44}$$

donde $f_c(L)$ representa una estimación a tamaño finito de f_c, qué típicamente se obtiene como la posición del máximo en la susceptibilidad o en el tamaño de la segunda componente.

En percolación aleatoria puede demostrarse que $\lambda = \nu$ [Sta87]. Es decir, el corrimiento del punto crítico está vinculado a la divergencia de la longitud de correlación. Si bien esto es válido en otros sistemas, en general el exponente λ difiere de ν. Por ejemplo, en algunos modelos de percolación explosiva los sistemas de tamaño pequeño presentan diferencias entre ambos exponentes [LÖ12; Gra+11]. A partir de la Ecuación (2.44) es posible estimar tanto el exponente λ como el umbral de percolación f_c. Sin embargo, este método sólo es aconsejable cuando es posible simular sistemas de gran tamaño. En caso contrario, es posible que introduzca errores sistemáticos [Bas+14].

Métodos de escaleo basados en estadísticas maximales

En un artículo publicado recientemente, Fan, et al. proponen un nuevo método para analizar procesos de percolación generalizados, el cual se basa en el escaleo del salto más alto en el parámetro de orden durante el proceso de percolación [Fan+20]. Si bien los autores abordan procesos de percolación de enlaces, podemos aplicar el análisis a percolación de nodos.

Como primera instancia, consideremos que nuestro proceso de percolación se repetirá muchas veces. Esto puede suceder de dos maneras distintas. O bien el proceso de percolación tiene alguna componente aleatoria que cambia de iteración en iteración, como es el caso de percolación clásica, o bien el proceso es determinista, pero contamos con un ensamble de redes con las mismas características estadísticas. En cualquiera de los dos casos, podemos definir observables para cada uno de los procesos individuales y luego considerar promedios y otras variables estadísticas.

Definimos para ello la cantidad

$$\Delta^{(i)} = \frac{1}{L^d} \operatorname{máx} \left[N_1^{(i)}(t+1) - N_1^{(i)}(t) \right], \tag{2.45}$$

donde $N_1^{(i)}(t)$ es el tamaño de la mayor componente luego de remover t nodos de la red en el i-ésimo proceso de percolación. Como se deduce de la definición, $\Delta^{(i)}$ es el mayor salto que se produce en la componente gigante en la iteración i del proceso. En adelante, nos referiremos a esta variable como *salto máximo*, o simplemente *salto*.

Definimos también $t_\Delta^{(i)}$ como la cantidad de nodos removidos al instante en que se produce el salto máximo, y $f_\Delta^{(i)} = t_\Delta^{(i)}/L^d$ como la posición del

salto máximo. Como veremos más adelante, este observable puede servir como estimador del punto de percolación para sistemas de tamaño finito. En general, su valor difiere del de otros estimadores, como por ejemplo la posición del pico de la susceptibilidad, pero estas diferencias tienen a disminuir en la medida que el tamaño del sistema aumenta. El tercer observable que definimos es el tamaño absoluto de la mayor componente en la posición del salto, es decir, $N^{(i)}_{1,\Delta}(L) \equiv N^{(i)}_1(t^{(i)}_\Delta(L))$.

Para cada una de las cantidades anteriores podemos definir un promedio de ensamble (o iteraciones) y su respectiva desviación estándar. Por ejemplo, para el salto máximo tenemos $\Delta(L) \equiv \langle \Delta^{(i)}(L) \rangle$ y $\chi_\Delta \equiv \langle (\Delta^{(i)}(L))^2 \rangle - \langle \Delta^{(i)}(L) \rangle^2$.

De acuerdo con [Fan+20], los promedios y fluctuaciones de los observables anteriormente definidos presentan un escaleo de la forma

$$\begin{aligned} \Delta(L) &\sim L^{-\beta/\nu}, & \chi_\Delta(L) &\sim L^{-\beta/\nu}, \\ f_\Delta(L) - f_\Delta(\infty) &\sim L^{-1/\nu_1}, & \chi_{f_\Delta}(L) &\sim L^{-1/\nu}, \\ N_{1,\Delta}(L) &\sim L^{-d_f}, & \chi_{N_{1,\Delta}}(L) &\sim L^{-d_f}. \end{aligned} \tag{2.46}$$

Las relaciones (2.46) nos brindan un método alternativo para determinar los exponentes ν, β y γ. Notemos además que, al igual que en la ecuación (2.44), el exponente asociado al corrimiento de la posición del salto ν_1 no necesariamente coincide con el de la longitud de correlación ν.

2.4.4. Ataques dirigidos basados en medidas de centralidad

En los ataques basados en medidas de centralidad los nodos son ordenados en una lista en orden decreciente de acuerdo con una medida de centralidad. Luego, los nodos son secuencialmente removidos siguiendo el orden establecido en esta lista. Si hay dos nodos que tienen la misma medida de centralidad estos se ordenan aleatoriamente. Existe una amplia variedad de medidas de centralidad que ha sido puesta a prueba para idear ataques sobre distintas redes. Uno de los análisis comparativos más completos entre métricas y tipos de redes puede verse en [Iye+13]. En particular, el grado de los nodos y betweenness se encuentran entre las medidas de centralidad más populares. Pero hay otras que también son muy utilizadas, como por ejemplo la centralidad por proximidad o *closeness*, la centralidad de autovector [New18] y la influencia colectiva [MM15]. Como regla general, cuando un nodo es removido las centralidades de los nodos que permanecen cambian. De manera que el ataque puede ser mejorado si se recalculan las listas de nodos a remover después de cada paso de remoción. Si la medida de centralidad usa información local, como es el caso de la centralidad por grado o la influencia colectiva, sólo una fracción de nodos va a cambiar su orden respecto de la lista original. Es decir, la lista se mantiene prácticamente invariante después de que algunos nodos son removidos. Por el contrario, los ataques por centralidad de betweenness o de autovectores utilizan medidas

globales y la remoción de un solo nodo puede afectar significativamente el ordenamiento de la lista. Es importante destacar que los ataques recalculados actualizan la información de la red a medida que avanza el ataque, por lo que en general son más eficientes que los ataque iniciales [Iye+13], aunque existen situaciones en las cuales el recalculado da lugar a un ataque menos efectivo [Hol+02].

Capítulo 3

Detección de comunidades en redes de jugadores de ajedrez

3.1. Resumen

En este capítulo utilizamos una de las bases de ajedrez mas extensas del mundo para construir dos redes de jugadores de ajedrez. Una de estas redes incluye juegos que tuvieron lugar sobre tablero físico (OTB) y la otra contiene partidas que se jugaron en portales de internet (Portal). Estudiamos las métricas y características topológicas mas importantes de estas dos redes, tales como su distribución de grado, existencia de correlaciones de grado, transitividad y asortatividad. Utilizamos además distintos algoritmos para analizar la estructura en comunidades de las redes, incorporando al análisis estructural el nivel de juego de los jugadores como un atributo asociado a los nodos. Aunque las dos redes son topológicamente diferentes, observamos que los jugadores se agrupan en comunidades acordes a su nivel de juego y que además emerge una estructura de *club de ricos*, constituida por los jugadores mas destacados.

3.2. El ajedrez como sistema complejo

El ajedrez es un juego emblemático que se destaca por su antigüedad, popularidad y complejidad. Se ha empleado para estudiar el comportamiento humano desde la perspectiva de muchas disciplinas, que van desde la evaluación de aspectos cognitivos, hasta situaciones que tienen que ver con la innovación y la toma de decisiones. Dado que una extensa recopilación de partidas ajedrez jugadas se encuentra disponible, es posible realizar análisis estadísticos detallados y significativos de este deporte.

El desarrollo de este juego esta íntimamente ligado al desarrollo reciente de nuestra civilización y, en la últimas décadas, también a la evolución de las computadoras [Pro12]. Los jugadores que se desempeñan en un alto nivel requieren de un duro entrenamiento mas allá de las condiciones naturales que posean, dada la enorme complejidad que caracteriza este deporte [Rib+13;

ASV16]. Esta complejidad, junto con la gran popularidad que tiene en todo el planeta, hace del ajedrez un medio adecuado para el estudio de varias áreas del comportamiento humano. Ha sido empleado, por ejemplo, para evaluar el desempeño cognitivo de jugadores profesionales y novatos [Dua+12], en el estudio de toma de decisiones [Sig+10; Leo+17] y para medir la relación entre expertiz y conocimiento [CG11; SG17]. También hay resultados muy interesantes respecto del comportamiento colectivo de un conjunto de jugadores [BT09; Per+13; SPB14; SPB16]. En uno de los trabajos más emblemáticos, Blasius y Tönjes [BT09] encontraron que la frecuencia con la que se utilizan las aperturas[1] sigue una ley de Zipf con exponente universal y, mediante un modelo constituido por un proceso multiplicativo, lograron reproducir sus observaciones. Mas aún, estudiando la dinámica de crecimiento del árbol de juego, se encontró que las leyes emergentes de Zipf y Heaps pueden explicarse mediante un proceso de tipo Yule-Simon anidado, es decir utilizando un proceso de crecimiento preferencial [Per+13] que es recurrente. Adicionalmente, se han reportado interesantes efectos de memoria en el registro temporal de juegos recopilados en grandes bases de datos [SPB14; SPB16].

Actualmente existen grandes comunidades de jugadores de ajedrez que involucran a jugadores de todos los niveles de juego y con diferentes grados de actividad. Las partidas pueden desarrollarse de manera presencial, sobre un tablero físico, o en diversos portales de internet. Estas extensas comunidades producen una gran cantidad de partidas que son apropiadamente almacenadas, proveyendo una fuente de datos muy útil para un análisis estadístico a gran escala. En estas bases de datos el nivel de juego está bien caracterizado estadísticamente mediante el sistema de puntuación *Elo*, introducido por el físico Arpad Elo [Gli95] y adoptado por la Federación Mundial de Ajedrez en el año 1970. Este sistema, que detallamos en el Apéndice A, no sólo permite ordenar a los jugadores de acuerdo a su rendimiento sino que además es un sistema con capacidad predictiva que permite anticipar el resultado de una partida de forma probabilística. De aquí que el sistema es útil para organizar los duelos en los torneos y para realizar un seguimiento de la evolución de los jugadores.

Como en casi toda actividad humana, es de esperar que los jugadores de ajedrez se agrupen entre ellos para formar redes estructuradas en comunidades. Hasta donde sabemos, las propiedades estadísticas de este tipo de redes, incluyendo la estructura de comunidades, no habían sido caracterizadas hasta que se publicó el trabajo asociado a esta tesis [Alm+17]. Cuando se considera una red de jugadores de ajedrez, el nivel de juego del jugador toma el rol de una información adicional sobre el nodo, lo que se conoce como *atributo* o *meta-dato*. Este nuevo elemento puede ofrecer información importante sobre las propiedades topológicas de la red. Por ejemplo, un comportamiento de tipo homofílico es de esperarse en estas redes, ya que es bien sabido [AE16] que los

[1] Cada partida de ajedrez consta de tres fases: apertura, medio juego y final. Se denomina apertura a la fase inicial del juego, en la que se procede a desarrollar las piezas desde sus posiciones iniciales.

jugadores de ajedrez tienden a jugar entre jugadores de su mismo nivel, es decir, existe una afinidad dada por el nivel de habilidad. Además, la relación entre la pericia de los jugadores y la estructura de la red puede proveer información acerca de los mecanismos que operan en la formación de comunidades en otros sistemas sociales.

El estudio de correlaciones entre la topología de una red social y los atributos de sus nodos es actualmente un tópico prominente en el área de las redes complejas [Leo+16; NC16; HPF16; HDF14]. La relación entre topología y atributos ya ha probado ser útil para detectar estratificaciones sociales [Leo+16], comunidades disortativas [NC16] y la inconsistencia de datos en el contexto de enlaces faltantes [HPF16]. Por otra parte, la inclusión de atributos también ha servido para señalar inconvenientes en los métodos tradicionales para la detección de comunidades [HDF14]. Sin embargo, los datos disponibles para estudiar este tipo de correlaciones son aún escasos y sólo existen unos pocos estudios de sistemas sociales que hayan sido analizados desde esta perspectiva. El propósito de este capítulo es describir las comunidades de jugadores de ajedrez en la escala mundial utilizando las herramientas para el análisis de las redes complejas, combinando los estudios de la estructura de estas redes con atributos de nodos. La idea es establecer similitudes y diferencias entre las redes de jugadores en tableros físicos y en los portales de Internet.

Dividimos los resultados del trabajo en cuatro secciones: en la Sección 3.3 introducimos las dos bases de datos sobre las cuales realizamos el análisis y mencionamos las principales características estadísticas de las mismas. En la Sección 3.4 describimos los métodos empleados para construir las redes a partir de los datos. La caracterización general de las redes, incluyendo un análisis de estructura y su relación con atributos de nodos, tiene lugar en la Sección 3.5. Completan el capítulo las Secciones 3.6, donde analizamos la emergencia de una estructura de tipo "club de ricos" que involucra a los jugadores expertos, y la Sección 3.7, en la cual estudiamos la estructura de comunidades de las redes y su relación con el nivel de juego de los jugadores.

3.3. Descripción de las bases de datos empleadas

Utilizamos dos bases de datos para construir las redes, ambas facilitadas en forma gratuita por la empresa *Opening Master* [Mas22], la cual aseguraba disponer de la base de datos de partidas de ajedrez mas grande del mundo en su momento. La primera base de datos, que denominaremos OTB, cuenta con unas 7.7 millones de partidas registradas, que fueron jugadas sobre tablero físico entre humanos. La segunda base de datos, que llamaremos Portal, cuenta con más de 15 millones de partidas desarrolladas en portales de Internet. La gran mayoría de estas partidas son entre humanos, pero también existen partidas entre humanos y computadoras. Los portales más representativos asociados a esta

base de datos son: chees.com, ICC-chessclub.com, Playchess.com. freechess.org y chesscube.com. Ambas bases de datos están en formato PGN, que es el formato más habitual para almacenar partidas de ajedrez, e incluyen información sobre campos tales como la denominación de los jugadores, su puntaje Elo, el resultado de la partida, la apertura utilizada y la secuencia de movimientos. La descripción detallada del formato PGN, junto con las herramientas utilizadas para su procesamiento y posterior análisis están detalladas en el Apéndice B.

En primer lugar analizamos los datos estadísticos relacionados al número de partidas en los que participan los jugadores de las dos bases de datos. En la Figura 3.1(a) mostramos el histograma normalizado para la frecuencia con que un jugador de la base de datos juega un número dado de partidas. Es decir, si tomamos un jugador al azar los valores de esta distribución nos dan la probabilidad de que ese jugador haya sido partícipe en un dado número de partidas. Se puede ver que en los dos casos, OTB y Portal, se trata de distribuciones sesgadas que en una escala log-log se asemejan a una ley de potencias. Esto quiere decir que hay muchos jugadores que han participado en un número pequeño de partidas y unos pocos que son muy activos. La linea punteada de la figura, que mostramos a modo de comparación, corresponde a una ley de potencia con exponente $-1{,}5$. En el caso de OTB, vemos una caída o *cutoff* exponencial hacia la derecha de la distribución. Caídas como estas son típicas en variables que se distribuyen de manera muy sesgada y pueden deberse a varios motivos. El principal de ellos es que, independientemente de qué tan grande sea la base de datos, la cantidad de datos disponibles es siempre limitada, con lo cual los eventos extremos (en este caso, los jugadores con mayor actividad) están en mayor o menor medida subrepresentados. En este caso, además, podemos esperar que esa caída persista, debido a que existe un límite para la cantidad de partidas que un jugador puede jugar a lo largo de su vida. La distribución de Portal se asemeja mucho a OTB, pero presenta su cutoff aproximadamente un orden de magnitud más adelante. Además, la distribución se aplana ligeramente al final, indicando la presencia de valores atípicos. La extensión del régimen de ley de potencias se debe a que esta base posee más partidas que OTB, mientras que los valores atípicos aparecen por la existencia de *bots* (jugadores virtuales constituidos por motores de ajedrez).

Por otra parte, en la Figura 3.1(b), mostramos un diagrama de rango-abundancia, en donde las abscisas representan a los jugadores, ordenados de manera descendiente de acuerdo al número de partidas jugadas, y las ordenadas la cantidad de partidas jugadas por cada jugador. Esta distribución representa la probabilidad de que una partida de la base de datos haya sido jugada por un jugador de un dado rango. Mediante este gráfico podemos apreciar nuevamente la naturaleza heterogénea de la actividad de los jugadores, es decir, que la gran mayoría de las partidas fue jugada por un número reducido de jugadores.

El siguiente análisis consistió en estudiar la distribución de puntajes Elo para cada jugador. Como mencionamos en el Apéndice A, el Elo es una cantidad dinámica que se actualiza después de que cada jugador termina una partida (o

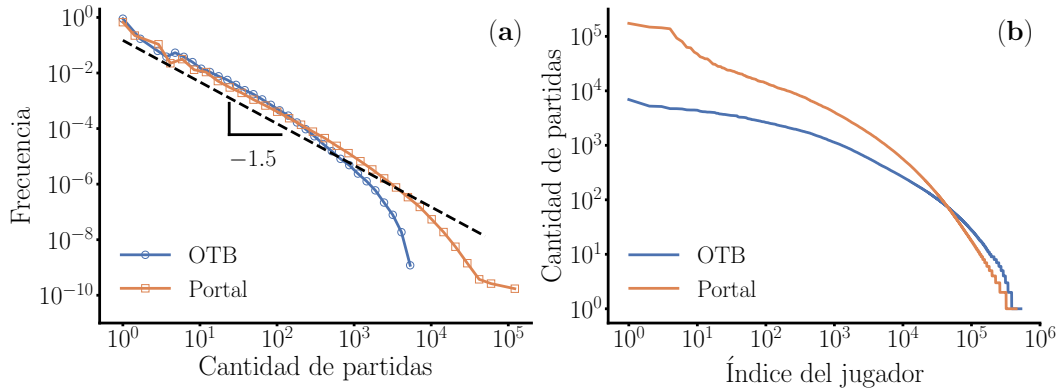

Figura 3.1: **(a)** Histograma normalizado para la distribución de la cantidad de partidas jugadas por jugador. Las distribuciones asociadas a ambas bases de datos (OTB y Portal) tienen un comportamiento similar, que puede ser aproximado por una ley de potencias con un decaimiento exponencial sobre el final. La mayor cantidad de datos de Portal, junto con la presencia de bots, hace que la distribución se estire un orden de magnitud más que la de OTB, y que presente un pequeño *plateau* hacia el final. La línea punteada sirve de referencia y representa una ley de potencia con exponente $-1{,}5$. **(b)** Diagrama de rango-abundancia para la cantidad de partidas por jugador.

un conjunto de partidas en el caso de torneos). Es decir, cada jugador tiene asociada una serie temporal con los valores de Elo obtenidos a lo largo de su trayectoria. Si bien no forma parte del presente trabajo, la evolución temporal del Elo es de por sí de gran interés. Algunos jugadores, sobre todos aquellos de mayor experiencia, mantienen su rango de Elo prácticamente constante, a excepción de pequeñas fluctuaciones. En cambio, los jugadores jóvenes que se destacan muestran un incremento sostenido de su Elo a lo largo del tiempo. No obstante, aquí decidimos simplificar el análisis, y en lugar de estudiar la serie temporal en su totalidad, analizamos sólo algunos estadísticos simples. En particular, calculamos el Elo medio de cada jugador $\langle \text{Elo} \rangle$ y su correspondiente desviación estándar σ_{Elo}. En la Figura 3.2(a) mostramos las distribuciones de Elo medio para el conjunto completo de jugadores, segregados según la base de datos a la cual pertenecen, mientras que en la Figura 3.2(b) graficamos las correspondientes distribuciones de las desviaciones estándar. En contraste con la actividad de los jugadores, las distribuciones de Elo medio son centradas. El valor medio del Elo para el caso de OTB es de 1915 ± 279, mientras que en Portal es 1767 ± 333. El hecho de que los jugadores de OTB tengan, en promedio, puntajes más bajos que en Portal es consecuencia del subregistro de partidas sobre tableros físicos de jugadores aficionados, a diferencia de los portales de internet, en donde el Elo se computa desde la primera partida jugada.

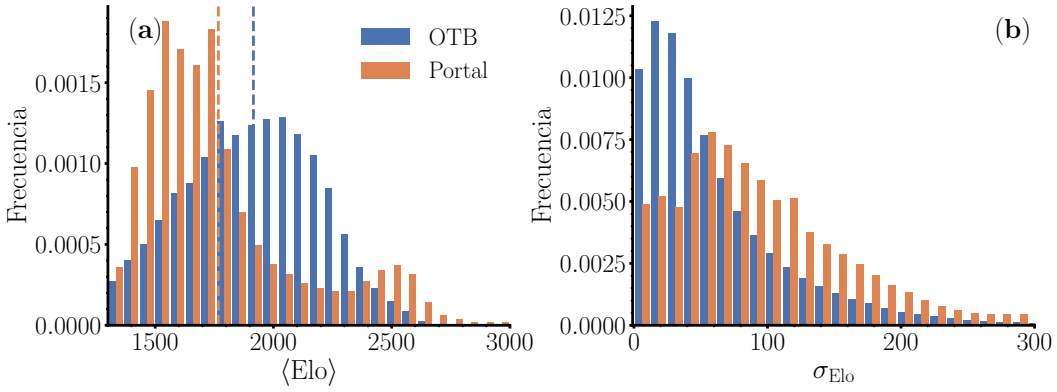

Figura 3.2: Análisis estadístico del Elo de los jugadores sobre tablero físico (OTB) y en portales de Internet. **(a)** Distribución de valores medios y **(b)** desviaciones estándar en el Elo de los jugadores. Las lineas verticales punteadas en **(a)** indican el valor medio de cada distribución (OTB y Portal).

En el juego del ajedrez, los jugadores prefieren enfrentar a otros jugadores con un nivel de juego similar, es decir, que tengan un Elo parecido [AE16]. Una de las razones es meramente motivacional; los jugadores no están interesados en participar de juegos donde sus chances de ganar o perder sean muy elevadas, o sea, donde se pueda anticipar el resultado. La otra razón está relacionada con el modo en que el Elo se actualiza después de cada partida. Si un jugador de alto Elo le gana una partida a un jugador de Elo considerablemente inferior, su ganancia en Elo es bastante reducida. Una derrota, por el contrario, conlleva una pérdida de Elo significativa (ver Apéndice A). Es por esto que los jugadores mejor puntuados evitan enfrentarse a jugadores de nivel muy inferior. Este comportamiento puede visualizarse a partir de los datos empleando un gráfico de densidad en dos dimensiones, donde se grafica para cada partida el Elo del jugador que utiliza las piezas blancas *versus* el Elo del jugador que emplea las piezas negras. En los paneles (a) y (b) de la Figura 3.3 presentamos estos gráficos, discriminados según la base de datos. De estas figuras se desprende que la mayoría de las partidas se desarrollan entre jugadores cuya diferencia de Elo, ΔE, no excede los 300 puntos. De hecho, la distribución de estas diferencia de Elo está bien descripta por una función exponencial, como se puede ver en los paneles (c) y (d) de la figura. El valor medio y la desviación estándar para estas diferencias son, respectivamente, $\langle \Delta E \rangle = 156$ y $\sigma_{\Delta E} = 126$ para OTB, mientras que para el Portal se obtiene $\langle \Delta E \rangle = 146$ y $\sigma_{\Delta E} = 159$. Además, podemos observar en los paneles superiores de la Figura 3.2 que si nos movemos hacia los Elo mayores, la dispersión entre el Elo de los jugadores se angosta, lo que implica que los jugadores de gran nivel tienden a formar comunidades cerradas.

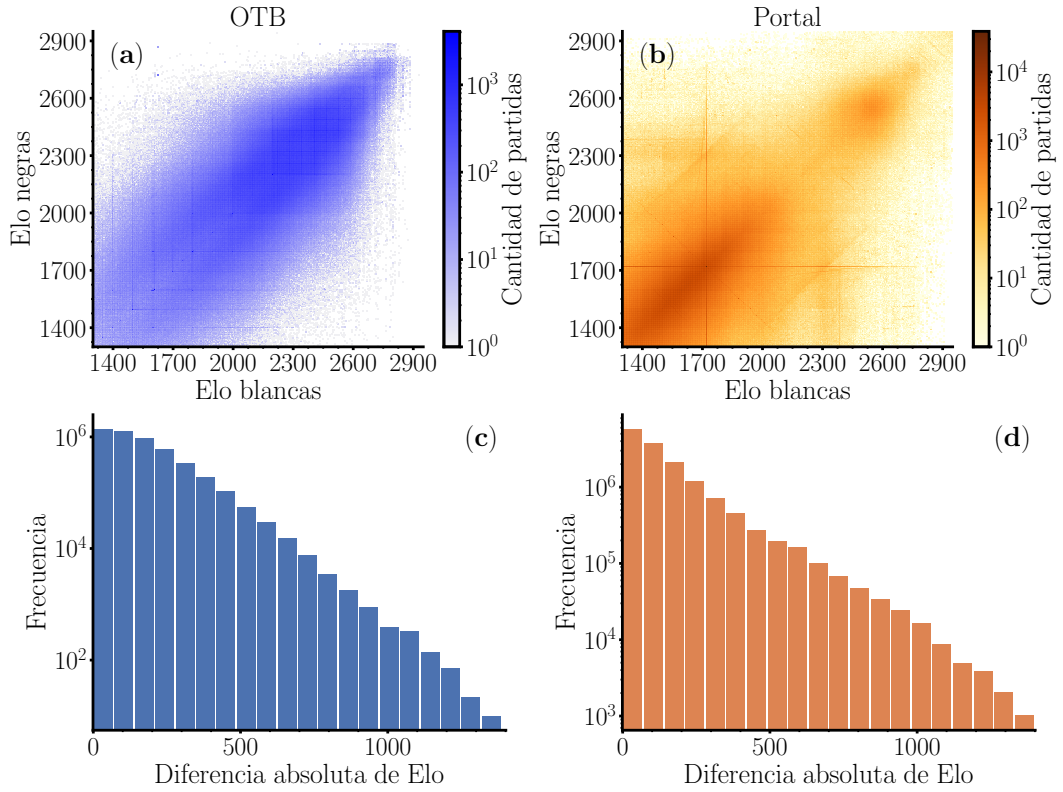

Figura 3.3: **(a,b)** Histograma bidimensional para los valores de Elo de los jugadores enfrentados al inicio de cada partida en cada base de datos. Las diferencias absolutas entre blancas y negras rara vez supera los 300 puntos, y se reduce a medida que aumenta el nivel de los jugadores. Nótese que las escalas de colores utilizadas son logarítmicas. **(c,d)** Histogramas en escala log-lineal para la distribución de las diferencias de Elo en cada base de datos. En ambos gráficos se observa un decrecimiento lineal, que corresponde a un decaimiento exponencial para la fracción de partidas en un función de las diferencias de Elo.

3.4. Construcción de las redes de jugadores

Con la información contenida en cada base de datos construimos dos redes, pesadas y no dirigidas. Como podemos observar en la Figura 3.4 cada nodo en la red representa un jugador individual y contiene además dos tipos de atributos, que son el nombre del jugador y su nivel de juego, expresado por su Elo medio. Las conexiones entre nodos se establecen siempre que los jugadores hayan disputado al menos una partida, y el peso de cada conexión corresponde a la cantidad de partidas jugadas por los dos jugadores que conecta. Además de estas redes, consideramos también las versiones no pesadas de las mismas. Generamos además versiones aleatorizadas a partir de permutaciones aleatorias en las conexiones las cuales fueron empleadas como modelos nulos.

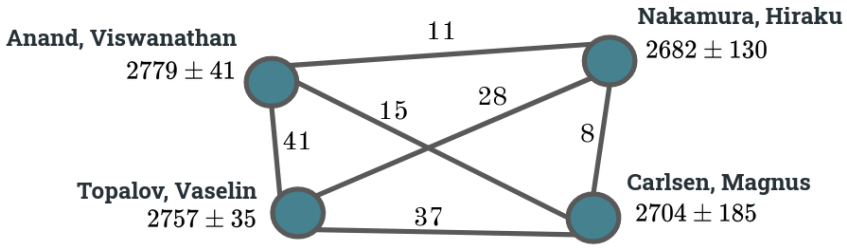

Figura 3.4: Representación de una subred compuesta de cuatro jugadores de alto rango de la base de datos OTB. Cada nodo representa a un jugador y contiene como atributo el nombre del jugador y su Elo medio (con la correspondiente desviación estándar). Si dos jugadores se ha enfrentado al menos una vez, estos se encuentran ligados por una conexión, cuyo peso corresponde al número de enfrentamientos.

3.5. Caracterización general de las redes de jugadores

Como es frecuente en el estudio de las redes complejas, comenzamos nuestra caracterización de las redes calculando sus distribuciones de grado. Consideramos para ello las versiones no pesadas de las redes, ignorando, por ahora, los atributos asociados a los enlaces y los nodos. Como mostramos en la Figura 3.5, las distribuciones de grado de ambas redes (OTB y Portal) son heterogéneas, con un sesgo hacia la derecha. Si bien hay similitudes entre las dos distribuciones, también se aprecian algunas diferencias. Mientras que la distribución OTB es cóncava en un gráfico con escala log-log, la distribución de grado de la red asociada a Portal se ajusta bien con una función de tipo ley de potencia con un exponente $\sim 1{,}5$ en un rango de 3 órdenes de magnitud, luego de lo cual presenta un decaimiento exponencial. El comportamiento observado (ley de potencias con decaimiento exponencial hacia los valores máximos) es un comportamiento típico observado en numerosas redes sociales [Bar99; New01] y actividades humanas [New05].

La similitud entre las distribuciones de grado y las correspondientes distribuciones de cantidad de partidas por jugador estudiadas en la Sección 3.3 (Figura 3.1) no es casualidad, ya que es esperable que un jugador que haya jugado muchas partidas haya enfrentado una mayor cantidad de oponentes que un jugador menos activo. De hecho, si consideramos las versiones pesadas de las redes, la fuerza de cada nodo (suma de los pesos de las conexiones con sus vecinos [Yoo+01]) representa la cantidad de partidas jugadas, por lo que su distribución es equivalente a la de la Figura 3.1.

A partir de las distribuciones de grado calculamos algunas medidas estadísticas, como por ejemplo el grado medio $\langle k \rangle$ y el parámetro de heterogeneidad $\kappa = \langle k^2 \rangle / \langle k \rangle$. Estos valores, junto con las características generales como la

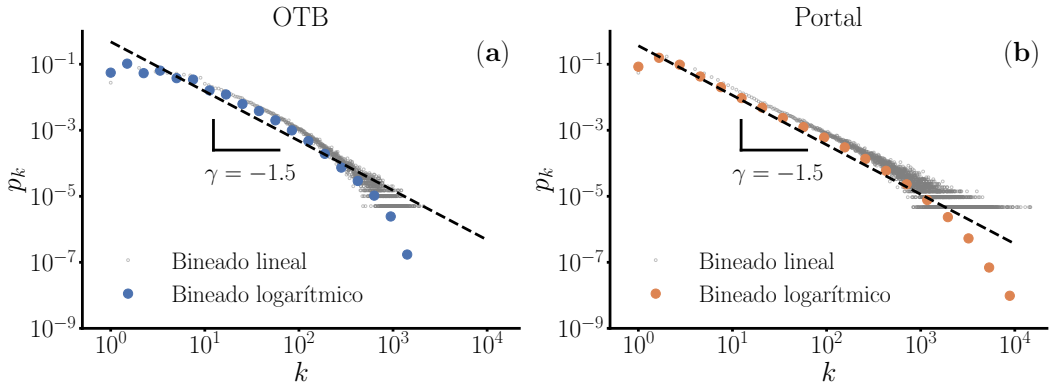

Figura 3.5: Distribución de grado de las redes. Los puntos grises representan un histograma con bineado lineal de tamaño 1 y los círculos de color, un histograma con bineado logarítmico. La línea punteada corresponde a una ley de potencias con exponente $-1{,}5$. **(a)** OTB y **(b)** Portal.

cantidad de nodos N y de enlaces M, y otras métricas que discutiremos más adelante, están sistematizadas en la Tabla 3.5. El valor acotado de $\langle k \rangle$ nos dice que las redes son esparsas, mientras que la relación $\kappa \gg \langle k \rangle$ resume la heterogeneidad observada en la distribución de grado.

Con la idea de explorar la existencia de correlaciones entre los grados de nodos vecinos calculamos la asortatividad por grado de las redes (ver Sección 2.1.3) mediante el coeficiente de correlación de Pearson r, definido por la Ecuación (2.16). Como vemos en la Tabla 3.5, la red OTB es asortativa mientras que la red de Portal es ligeramente disortativa. Como ha sido previamente reportado en la literatura [New02], muchas redes sociales con una cola larga en las distribuciones de grado son asortativas, como por ejemplo las redes de colaboraciones científicas y las redes de colaboración entre actores de cine [New03a]. El punto interesante en nuestro análisis es que la red Portal, siendo una red social, se aparta de este comportamiento y se asemeja más a una red tecnológica o biológica [New02].

De manera complementaria al coeficiente de Pearson, estudiamos las correlaciones analizando el grado medio de primeros vecinos $k_{nn}(k)$. Como mencionamos sobre el final de la Sección 2.1.3, un comportamiento creciente de k_{nn} con k indica correlaciones positivas entre grados. Es decir, una red con este comportamiento es de carácter asortativo. Si observamos el panel (a) de la Figura 3.6, podemos apreciar que OTB cumple con esta característica. En contraste, el panel (b) de la figura nos muestra que la red Portal tiene correlaciones negativas. Antes de sacar alguna conclusión sobre estas diferencias, conviene mencionar que las distribuciones heterogéneas de grado pueden imponer restricciones topológicas que dan lugar a correlaciones [BPV04; MAA02]. Estas correlaciones son inherentes a la red y no necesariamente responden a los

	OTB	OTB CM	Portal	Portal CM
N	$2{,}0 \times 10^5$	–	$3{,}9 \times 10^5$	–
M	$2{,}1 \times 10^6$	–	$7{,}6 \times 10^6$	–
$\langle k \rangle$	39	–	71	–
κ	235	–	1200	–
C	0,106	0,007	0,086	0,055
C_{WS}	0,188	0,007	0,179	0,096
r (grado)	0,359	−0,001	−0,118	−0,095
r (Elo)	0,693	−0,004	0,557	0,000
Q	0,69	0,16	0,42	0,10
μ	0,24	0,88	0,40	0,93
N_C	3604	1416	12734	3861

Cuadro 3.1: Medidas estadísticas globales calculadas sobre las versiones no pesadas de las redes estudiadas, y sobre redes aleatorias asociadas generadas por el modelo de configuración (CM). N: cantidad de nodos, M: cantidad de enlaces, $\langle k \rangle$: grado medio κ: parámetro de heterogeneidad, C: coeficiente de *clustering* (Newman) y C_{WS}: coeficiente de *clustering* (Watts-Strogatz), r: parámetro de asortatividad (por grado y por Elo), Q: modularidad, μ: parámetro de mezcla y N_C: número de comunidades. Los métricas Q, μ y N_C fueron calculados sobre una realización del algoritmo de Leiden (ver Sección 2.3).

principios evolutivos con los que la red fue formada. Por este motivo, resulta conveniente comparar los resultados obtenidos en las redes analizadas con los mismos resultados aplicados sobre un adecuado modelo nulo. La elección más apropiada de modelo nulo resulta ser la de una red que tenga la misma distribución de grado que la red original, pero que en lo demás sea aleatoria. En este caso, empleamos dos modelos nulos: el modelo de configuración (Sección 2.2.2 y el intercambio de conexiones, o recableado (Sección 2.2.3).

Volviendo al panel (a), podemos observar que las redes aleatorias construidas con la distribución de grado de OTB tienen una muy baja correlación de grado. De hecho, el valor de k_{nn} no depende significativamente de k y se aproxima al resultado esperado para una red sin correlaciones, cuyo valor es $k_{nn}^{\text{unc}} = \langle k^2 \rangle / \langle k \rangle$ [VBB08]. También verificamos que no existen diferencias entre los dos modelos nulos empleados. Por otra parte, las redes aleatorias construidas con la distribución de grado de Portal mantienen una clara correlación negativa similar a la de la red original. Con esto podemos asegurar que la asortatividad de OTB es genuina, en el sentido en que es consecuencia de la dinámica evolutiva de la red, mientras que el comportamiento disortativo de Portal se debe principalmente a sus restricciones estructurales, particularmente a la gran heterogeneidad de su distribución de grado. El mismo efecto se observa al comparar el valor del coeficiente de Pearson de cada red con su respectiva red aleatorizada (ver Tabla 3.5).

Además de la asortatividad por grado, calculamos la asortatividad por Elo,

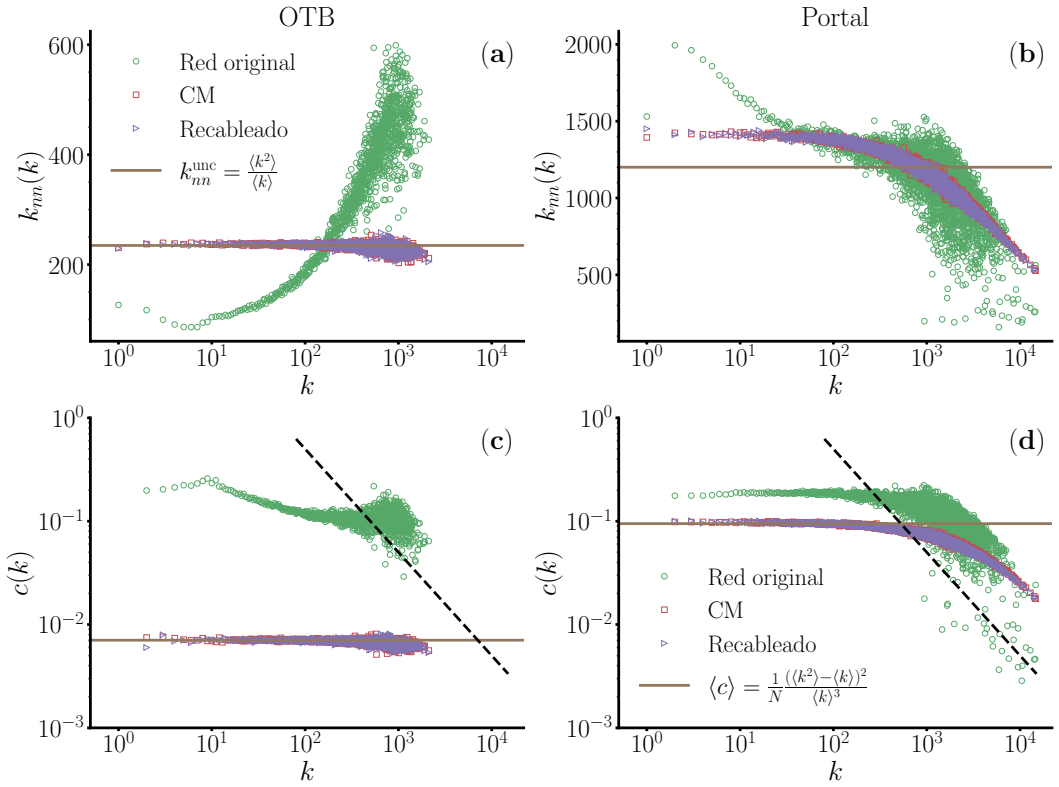

Figura 3.6: Propiedades estructurales de las redes OTB y Portal. **(a,b)** grado promedio k_{nn} de los vecinos a un nodo de grado k. Se observa en las redes originales que OTB es asortativa mientras que Portal es disortativa. Cuando se aleatorizan las redes (utilizando el modelo de configuración (CM) y el intercambio de dos conexiones (Recableado), descriptos en la Sección 2.2) las correlaciones desaparecen en OTB y k_{nn} toma el valor constante correspondiente a una red aleatoria $\langle k^2 \rangle / \langle k \rangle$ (linea horizontal). En cambio, en Portal no es posible eliminar completamente las correlaciones. **(c,d)** Coeficiente de clustering en función del grado de los nodos para ambas redes. Las lineas horizontales indican los valores correspondientes a redes sin no correlacionadas con la misma distribución de grado. La línea punteada representa la dependencia $c(k) \sim k^{-1}$ que se espera para una red jerárquica. Nuevamente, la aleatorización de las redes elimina las correlaciones en OTB mientras que en Portal permanecen, aunque en el rango de valores de k menores a 10^3 se aproximan al valor esperado en una red sin correlaciones de grado.

utilizando la métrica introducida en la Sección 2.1.3 para mezcla asortativa por atributos ordinales. A diferencia de lo que observamos para la asortatividad por grado, en este caso las dos redes son asortativas (ver valores en la Tabla 3.5). De hecho, en ambas redes los coeficientes de asortatividad son significativamente altos cuando se los compara con los valores asociados a los modelos nulos.

Como veremos más adelante, este hecho es un primer indicador de que la dinámica evolutiva de las redes se rige principalmente por el nivel de juego de los jugadores.

Otro elemento que estudiamos en nuestras redes fue la transitividad. Para cuantificar este fenómeno, calculamos los dos coeficientes globales introducidos en la Sección 2.1.2: el coeficiente de clustering de Newman (Ecuación (2.4)) y el coeficiente de Watts-Strogatz (Ecuación (2.6)). Como podemos apreciar en la Tabla 3.5, el valor de C es similar para ambas redes, y lo mismo sucede con C_{WS}. Si, en cambio, comparamos las dos métricas para una misma red, verificamos $C_{WS} > C$, desigualad frecuentemente observada en redes con distribución de grado heterogénea (ver discusión en la Sección 2.1.2). Por último, podemos asegurar que los valores de transitividad obtenidos son estadísticamente significativos comparándolos con los valores calculados sobre redes aleatorizadas. Para ambas redes, sus correspondientes redes aleatorias muestran valores menores de transitividad. Sin embargo, notamos que la diferencia es mayor en OTB que en Portal.

Calculamos, además, el coeficiente de clustering $c(k)$ en función del grado k (Ecuación (2.7)). En los paneles (c-d) de la Figura 3.5 mostramos que en las dos redes $c(k)$ es una función decreciente, pero que decrece de manera más notoria en la red Portal que en la OTB. Este comportamiento, que ha sido previamente observado en redes tales como las de comercio mundial [Fag07] y las metabólicas [Rav+02], se reporta como un buen indicador de la existencia de una estructura jerárquica en las redes [Rav+02; RB03]. En particular, el modelo original propuesto por Ravasz predice una dependencia de tipo ley de potencia para el coeficiente de clustering de la forma $c(k) \sim k^{-1}$ cuando la red es jerárquica [Rav04]. Las lineas punteadas en las gráficas indican esta relación.

Como en las redes disortativas el coeficiente de clustering de los nodos muy conectados está acotado superiormente [SB05], las redes disortativas están por regla general menos clusterizadas que la asortativas [Fos+11]. Esto es consistente con nuestros resultados (OTB es asortativa y está mas clusterizada que Portal, que es disortativa), lo que muestra que en la red Portal existe una estructura jerárquica que es más marcada que en OTB. Después de la aleatorización, las correlaciones en $c(k)$ desaparecen completamente en OTB, sin importar el algoritmo de aleatorización que se emplee. En Portal, los algoritmos de aleatorización que empleamos reducen sustancialmente la dispersión y las correlaciones, pero estas últimas no son eliminadas por completo. A modo de referencia, indicamos en los paneles el valor $\langle c \rangle = (\langle k^2 \rangle - \langle k \rangle)^2/(N\langle k \rangle^3)$, el cual corresponde al valor esperado para el coeficiente de clustering en el modelo de configuración [New18].

3.6. El "club de ricos"

En este apartado analizamos el fenómeno "club de ricos" en las redes de jugadores. Los resultados de este análisis se pueden ver en la Figura 3.7. En

el panel (a) de la figura mostramos los coeficientes $\phi(k)$ (Ecuación (2.21)). La red OTB muestra un comportamiento estrictamente creciente para este coeficiente. Por otro lado, Portal es creciente en un amplio rango de k, pero decrece en una región acotada ubicada en valores altos de k. Como discutimos en la Sección 2.1.5, un valor positivo de $\phi(k)$ no es suficiente para asegurar la existencia de un club de ricos genuino, sino que es necesario comparar este resultado con el de una red aleatorizada con la misma distribución de grado. Las líneas discontinuas de la figura muestran precisamente este cálculo, realizado sobre redes aleatorizadas utilizando el método de intercambio de conexiones (ver 2.2.3). Podemos verificar que el comportamiento creciente de esta métrica persiste aún en las redes aleatorizadas. De hecho, en el caso de Portal es incluso más significativo. En el panel (b) de la figura graficamos el cociente $\rho(k)$, correspondiente al coeficiente de club de ricos normalizado (Ecuación 2.22). Observamos ahora una clara diferencia entre ambas redes. Por un lado, la existencia de un club de ricos se verifica para OTB, por lo que la red, en este sentido, se comporta de manera similar a otras redes sociales. Curiosamente, Portal no sólo carece de club de ricos, sino que además los nodos de alto grado tienden a estar menos conectados entre sí que en su correspondiente red aleatoria. Una vez más, vemos en Portal características que la asemejan más a una red tecnológica que a una red social.

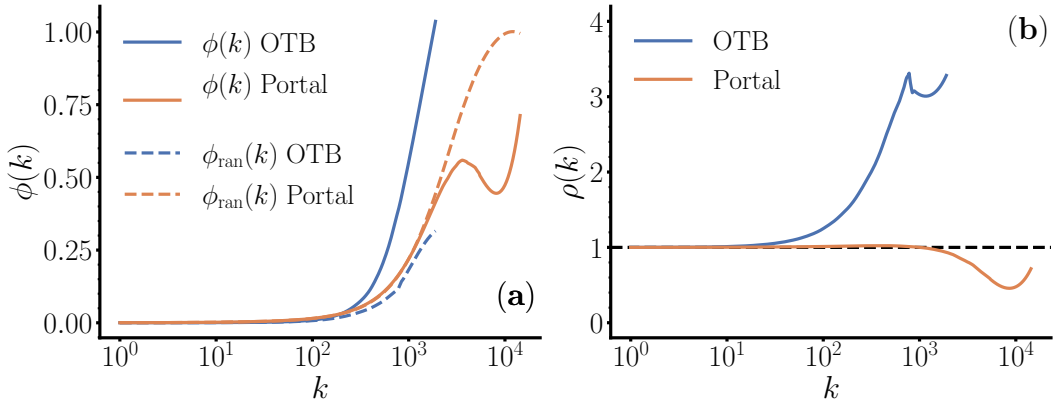

Figura 3.7: **(a)** Coeficiente de club de ricos sin normalizar para las redes originales OTB y Portal ($\phi(k)$) y las correspondientes aleatorizaciones ($\phi_{ran}(k)$). En todos los casos se evidencia el fenómeno. **(b)** Coeficiente de club de ricos normalizado $\rho(k)$. Podemos observar que el fenómeno club de ricos se observa sólo en la red OTB.

En su propuesta original, el club de ricos está definido en términos de una propiedad estructural de la red (el grado). Sin embargo, nada nos impide extender la definición reemplazando esta propiedad por algún atributo de los nodos que represente, en un contexto asociado a la red, algún tipo de riqueza. Por ejemplo, podríamos pensar en una red social cuyos nodos sean personas y

tomar el nivel de ingresos de cada persona como atributo, con el fin de evaluar si las personas pertenecientes a estratos sociales de mayores ingresos (literalmente, los ricos) tienden a conectarse preferentemente entre sí. En el trabajo [Alm+17], asociado a esta tesis, propusimos una extensión del concepto de club de ricos en esta dirección, que fue luego ampliada por Cinelli en [Cin19].

En el contexto del ajedrez, el atributo más natural para asociar el concepto de riqueza es el Elo. En términos de este atributo, definimos entonces el coeficiente de club de ricos asociado como,

$$\phi(\text{Elo}) = \frac{2M_{>\text{Elo}}}{N_{>\text{Elo}}(N_{>\text{Elo}} - 1)}, \tag{3.1}$$

donde $M_{>\text{Elo}}$ y $N_{>\text{Elo}}$ son las conexiones remanentes después de que se remueven los nodos (y sus conexiones) correspondientes a jugadores con un Elo medio menor o igual a un dado valor. Definimos además $\phi_{ran}(\text{Elo})$ como el coeficiente calculado sobre una red aleatorizada utilizando el método de intercambio de conexiones y $\rho(\text{Elo}) = \phi(\text{Elo})/\phi_{ran}(\text{Elo})$, de manera análoga a las respectivas definiciones en términos del grado. Notemos que, al utilizar el método de intercambio de conexiones, la red aleatoria mantiene la correlación existente entre el grado y el atributo (Elo, en este caso) de cada nodo. Mantener esta correlación es necesaria, ya que de lo contrario obtendríamos un coeficiente de club de ricos normalizado trivialmente nulo.

Los resultados de este análisis están contenidos en la Figura 3.8. En el panel (a) de la figura podemos ver que ambas redes exhiben un coeficiente de club de ricos (sin normalizar) mayor a cero para los jugadores de más alto rango (Elo > 2300), siendo este más notorio en la red Portal que en OTB. En el panel (b) mostramos el coeficiente asociado a las redes aleatorizadas. En este caso vemos que, para OTB, existe un crecimiento de $\phi_{ran}(\text{Elo})$ con el Elo hasta un valor aproximado de Elo $= 2500$, luego de lo cual comienza a descender, aunque manteniendo siempre valores positivos. Por otra parte, el coeficiente asociado a Portal decrece en la región (Elo < 2300), luego de lo cual crece ligeramente. Cualitativamente, estas curvas pueden entenderse considerando la correlación existente entre el Elo y el grado k de los nodos en cada red. Como podemos ver en el panel (c), para OTB esta correlación es siempre positiva. Por este motivo, es esperable el fenómeno de club de ricos de Elo guarde similitud con el correspondiente para grado, tanto en la red original como en la red aleatorizada. En cambio, la correlación entre Elo y grado para Portal es ligeramente negativa excepto para los jugadores más activos, en donde se vuelve positiva. Por último, el panel (d) de la figura muestra el cociente $\rho(Elo)$ para cada una de las redes. En este caso, la existencia de un club de ricos genuino se vuelve evidente en ambas redes, por lo que podemos decir que existe, tanto para OTB como para Portal, una "oligarquía" compuesta por los jugadores de mayor nivel con una conectividad mayor a la esperada.

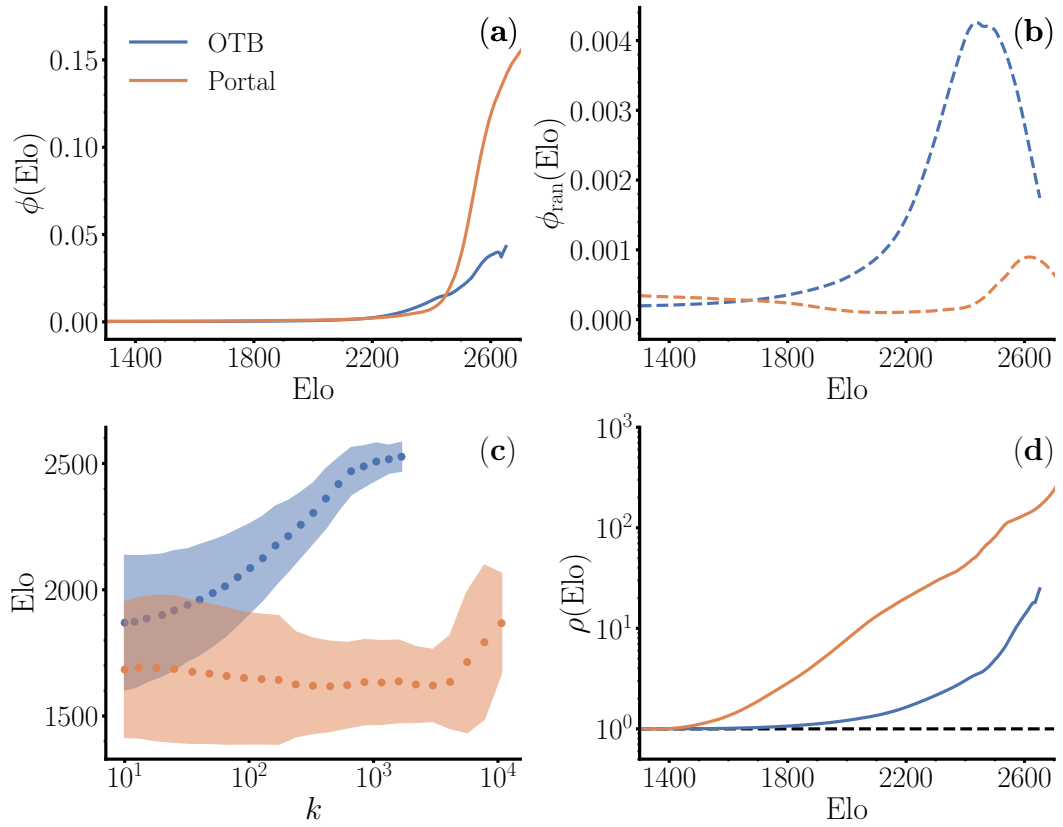

Figura 3.8: **(a)** Coeficiente de club de ricos sin normalizar definido en función del Elo para las redes originales OTB y Portal. **(b)** Mismo coeficiente calculado sobre redes aleatorizadas. **(c)** Correlación entre el Elo medio de los jugadores y su grado, analizado en los dos redes. En la red OTB se observa una correlación positiva, mientras que en Portal hay una ligera correlación negativa a excepción de los jugadores más activos, en donde el efecto se revierte. Cada punto representa el promedio entre jugadores con Elo similar, utilizando un bineado logarítmico, y el sombreado su correspondiente desviación estándar. **(d)** Coeficiente de club de ricos por Elo normalizado.

3.7. Estructura de comunidades

3.7.1. Comunidades basadas en la modularidad

Como mencionamos en la Sección 2.3, existe una amplia variedad de algoritmos desarrollados para detectar comunidades mediante la maximización de la modularidad. Sin embargo, muchos de estos algoritmos se vuelven impracticables en redes grandes debido a su alto costo computacional. Las redes aquí estudiadas tienen un tamaño del orden de $N \sim 10^5$, por lo que sólo podemos emplear algoritmos de complejidad lineal o casi lineal. Por ejemplo, el popular algoritmo de Newman-Girvan [NG04], basado en el cálculo de betweenness,

debe ser descartado ya que su complejidad computacional es $\mathcal{O}(N^3)$. Teniendo en cuenta esta restricción, decidimos emplear los algoritmos de Louvain y de Leiden, ambos introducidos en la Sección 2.3.1, cuya complejidad es $\mathcal{O}(N \log N)$.

Para caracterizar la estructura de comunidades evaluamos la modularidad Q, el parámetro de mezcla μ y la cantidad de comunidades N_c halladas por cada algoritmo. En la Tabla 3.5 mostramos los valores obtenidos para estas métricas empleando el algoritmo de Leiden, con el cual obtuvimos la mejor maximización de la modularidad. Encontramos que las comunidades se encuentran mejor definidas en la red OTB que en Portal, dado que los valores de Q en OTB son mayores a los de Portal mientras que los valores de μ son menores. En ambos casos se cumple además el criterio de Clauset, et al. ($Q > 0{,}30$) [CNM04], que sugiere que la estructura modular es significativa. Más aún, al aplicar los algoritmos de detección sobre las versiones aleatorizadas de las redes, podemos ver que Q disminuye y que μ aumenta notoriamente.

Una vez determinada la existencia de comunidades en las redes, continuamos estudiando la presencia de correlaciones entre el Elo de los jugadores y la pertenencia a las comunidades. Para llevar a cabo este punto calculamos el Elo medio del conjunto de jugadores que pertenece a cada comunidad y ordenamos las comunidades de acuerdo con este valor. Cada panel de la Figura 3.9 muestra los resultados correspondientes a una red en particular (OTB, Portal, y sus versiones aleatorizadas), empleando el algoritmo de Leiden, y son similares a los obtenidos empleando Louvain, como puede verse comparando con [Alm+17]. El tamaño de cada símbolo es proporcional al logaritmo del tamaño de la comunidad que representa, y las regiones sombreadas corresponden a la desviación estándar de Elo de cada comunidad. Para asegurar una estadística mínima en cada comunidad, sólo consideramos aquellas comunidades con más de 20 integrantes. Incluimos además, a modo de referencia, el valor medio del Elo calculado sobre todos los jugadores de la red (línea verde punteada). El panel (a) de la figura muestra el caso de la red OTB. Como podemos ver, existe una variación notoria entre el Elo medio de las primeras comunidades y el de las últimas. En particular, las primeras dos comunidades encontradas por el algoritmo tienen un Elo medio en torno a 2300, que corresponde a jugadores profesionales con nivel internacional. Luego de estas comunidades, el Elo medio cae a un valor cercano a 2000, correspondiente al nivel de torneos de alcance nacional (ver Tabla A.1). Comparemos ahora con el panel (b), donde mostramos los resultados correspondientes a la versión aleatorizada de OTB. Podemos ver ahora que no hay variación significativa de Elo entre las distintas comunidades. La cantidad de comunidades también varía, ya que para esta red, Leiden detecta pocas comunidades con más de 20 integrantes, y todas ellas tienen aproximadamente el mismo tamaño.

Al analizar la red Portal (panel (c)) podemos notar algunas diferencias con respecto a OTB. En primer lugar, las fluctuaciones de Elo dentro de cada comunidad son menores, por lo que podemos decir que el Elo medio de las

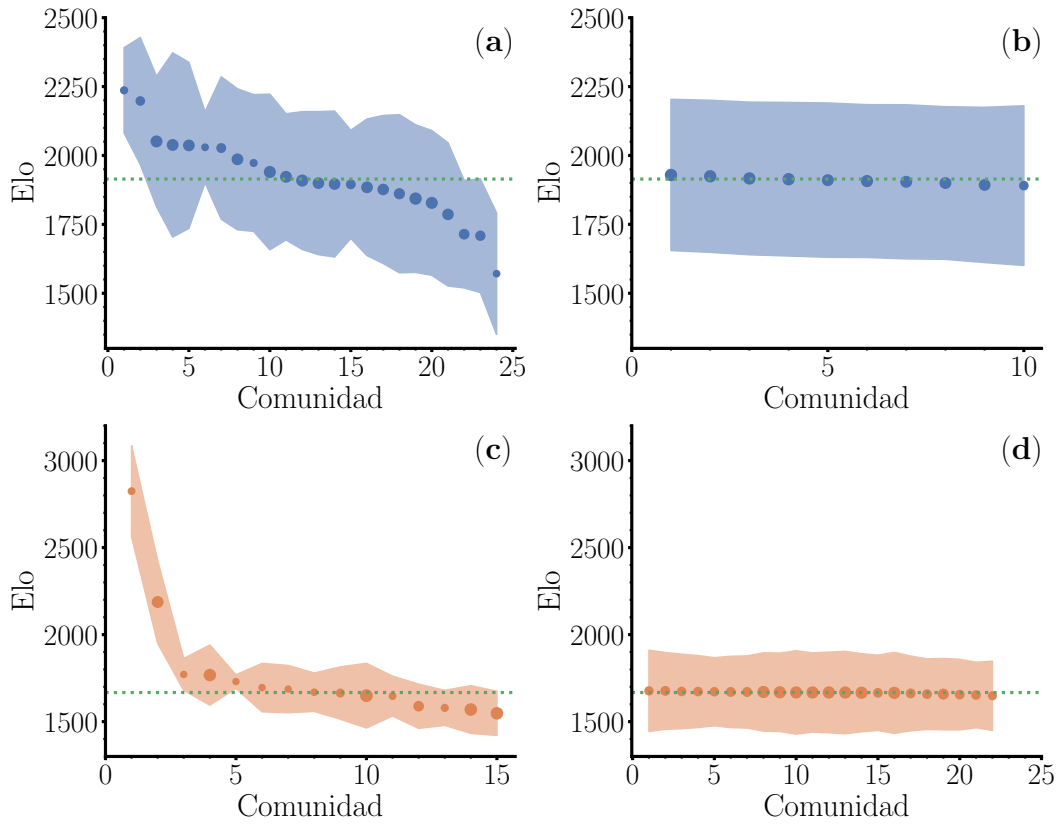

Figura 3.9: Correlaciones entre el Elo medio de un jugador y su pertenencia a una dada comunidad. Cada punto representa una comunidad, con tamaño de símbolo proporcional al logaritmo del tamaño de la comunidad (sólo se muestran comunidades con más de 20 integrantes). En todos los casos las comunidades están ordenadas de mayor a menor de acuerdo al valor medio de su Elo. Además, la región sombreada representa la desviación estándar en el Elo. La línea verde punteada corresponde al Elo medio calculado sobre todos los jugadores de la red. **(a)** OTB; **(b)** OTB aleatorizada; **(c)**; Portal **(d)** Portal aleatorizada. Observamos que en las dos redes originales las comunidades se pueden discriminar por su Elo medio, mientras que en las versiones aleatorias el Elo medio de todas las comunidades es muy similar. Además, el número de comunidades se reduce al aleatorizar (en OTB particularmente), y las fluctuaciones dentro de cada comunidad aumentan.

comunidades está mejor definido en esta red. Por otra parte, hay dos comunidades que destacan por tener un Elo considerablemente alto, diferenciándose del resto. Es decir, el algoritmo de Leiden identifica en este caso dos comunidades con jugadores expertos bien definidas. Al aleatorizar la red (panel (d)), vemos nuevamente un conjunto de comunidades que comparten el mismo valor aproximado de Elo, y fluctuaciones considerablemente mayores.

La correlación entre Elo y comunidades se puede visualizar de manera

alternativa calculando el coeficiente de variación del Elo (CV), esto es, el cociente entre las desviación estándar del Elo y su media. En la Figura 3.10 graficamos el CV para cada comunidad en función de su Elo medio. Podemos ver que la variabilidad del Elo es, en promedio, menor dentro de las comunidades que en las redes completas (línea negra discontinua). Mas aún, comparando estos resultados con aquellos obtenidos en los modelos nulos (regiones sombreadas en las figuras), vemos que que la diferencia es evidente. Expresado en otras palabras, el algoritmo de Leiden es capaz de discriminar eficientemente a los jugadores de acuerdo a su rendimiento.

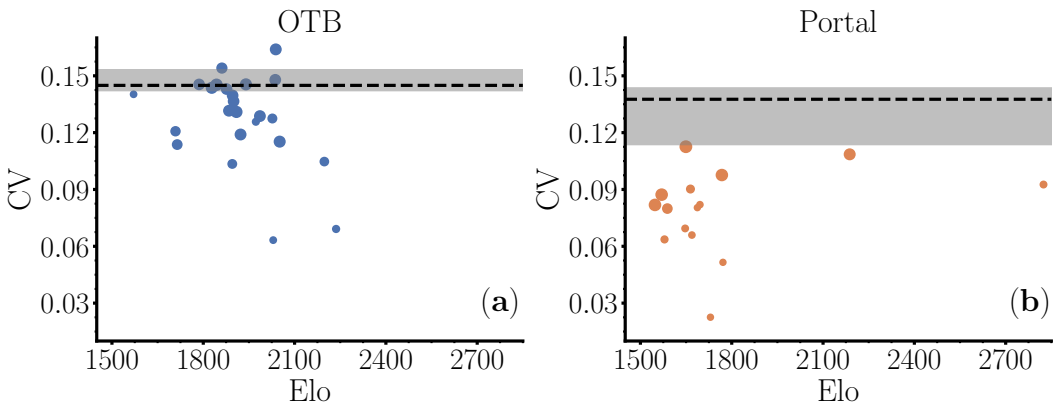

Figura 3.10: Coeficiente de variación en función del Elo medio de las comunidades. Las líneas discontinuas negras corresponden a los coeficientes de variación en el Elo calculados sobre todos los jugadores de cada red. Las regiones sombreadas representan el rango en el cual todos los CV de las comunidades de las redes aleatorizadas están incluidos. Muchas de las comunidades (particularmente de Portal) tienen un coeficiente de variación menor que el caso aleatorio.

3.7.2. Comunidades basadas en atributos de los nodos

Con la idea de detectar explícitamente la correlación entre comunidades y el Elo de los jugadores analizamos la estructura de las comunidades usando el algoritmo desarrollado por Newman y Clauset [NC16] (ver Sección 2.3.2). Continuando con la línea de análisis llevada a cabo en este capítulo, elegimos el Elo medio de los jugadores como atributo para incorporar al algoritmo. Como el algoritmo empleado sólo admite atributos categóricos, fue necesario definir categorías en función del Elo. Para ello, utilizamos categorías convencionales definidas por la Federación Internacional de Ajedrez (FIDE), definidas según rangos de Elo (Tabla A.1). En cuanto a la cantidad de comunidades, que en este algoritmo es un parámetro K fijado *a priori*, fue variada entre $K = 2$ y $K = 9$. Cualitativamente, los resultados no difieren al cambiar el número de comunidades, por lo que sólo mostramos los resultados correspondientes a dos

comunidades.

El análisis realizado está condensado en la Figura 3.11. Cada panel de la figura muestra, para cada categoría de atributo, la ocupación del total de nodos que pertenece a esa categoría en cada una de las dos comunidades determinadas por el algoritmo. Los paneles (a) y (b) corresponden, respectivamente, a las redes OTB y Portal. En ambos casos vemos una comunidad compuesta principalmente de jugadores aficionados (categorías "Clase A" e inferiores), y otra compuesta principalmente de jugadores profesionales (categoría "Experto" y superiores). A simple vista podemos ver que la división es más notoria en Portal que en OTB, lo cual puede verificarse calculando los valores de información mutua normalizada (NMI = 0,2191 para OTB y NMI = 0,4650 para Portal). Para descartar la posibilidad de que las comunidades encontradas fuesen un artefacto del algoritmo empleado, repetimos el proceso de detección sobre redes aleatorizadas. Utilizamos, para ello, tres formas diferentes de aleatorización. En primer lugar, tal como hicimos anteriormente, realizamos un intercambio de dos conexiones. En este esquema el algoritmo se vuele particularmente sensible a la secuencia de números aleatorios, donde encontramos que algunas veces las comunidades se correlacionan, de manera espuria, con el Elo de los jugadores. En otro de los modelos nulos realizamos una mezcla intercambiando el Elo de los jugadores. Este es el caso mostrado en los paneles (c) y (d) de la figura, donde puede verse que la correlación entre el Elo y las comunidades desaparece tanto para OTB como para Portal. Los valores de información mutua normalizada, en este caso, son NMI = 0,0003 para OTB y NMI = 0,0010 para Portal, dos órdenes de magnitud menores que para las redes originales. En el tercer modelo nulo realizamos una doble aleatorización, es decir después de intercambiar las conexiones aleatorizamos el Elo. Al igual que en la primera opción, el intercambio de conexiones vuelve inestable al algoritmo, por lo que decidimos descartarla.

3.8. Conclusiones del capítulo

Como un complemento a los juegos tradicionales sobre el tablero, los portales de Internet han introducido un nuevo modo de jugar al ajedrez, creando en el proceso dos conjuntos paralelos de jugadores. Si bien el juego es el mismo, cada conjunto tiene modos particulares de evolución que lo lleva a presentar características propias. Al analizar la base de datos OTB, observamos que los juegos recopilados se enfocan principalmente en jugadores profesionales, con escasos jugadores aficionados. Esto es de algún modo esperable ya que la actividad de los jugadores aficionados se desarrolla en contextos informales en donde los juegos no son registrados. Una de las consecuencias de este sesgo es que la base OTB tiene jugadores con Elos mayores que Portal y que además en OTB el Elo evoluciona mas lentamente. En cambio, en Internet este aspecto es completamente diferente ya que todos los juegos son almacenados, independientemente del Elo de los jugadores. En consecuencia, los jugadores

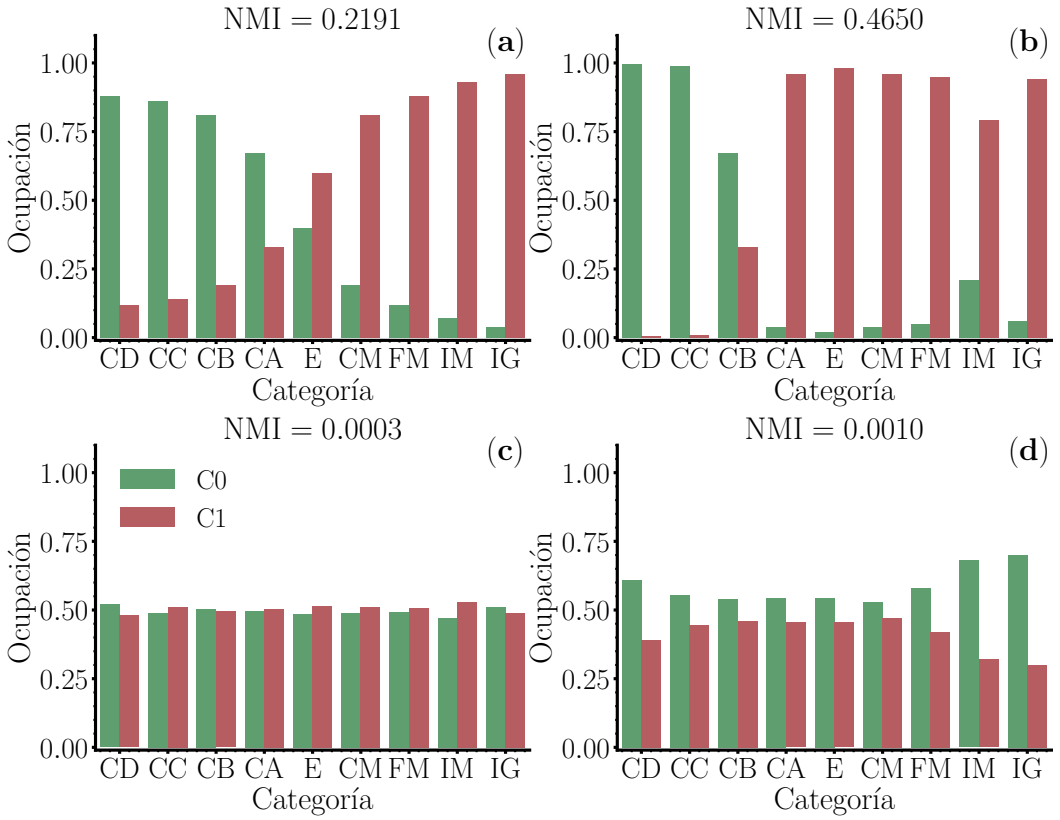

Figura 3.11: Distribución de Elo de cada comunidad detectada por el algoritmo de Newman y Clauset fijando $K = 2$. El color de las barras identifica a la comunidad y las letras del eje de las abscisas representan las categorías definidas en cada rango de Elo, ordenadas de menor a mayor rango (para una definición precisa de las categorías ver el Apéndice A). Los valores correspondientes de la información mutua normalizadas (NMI) están indicados en los títulos de las figuras. **(a-b)** Redes originales. **(c-d)** Modelo nulo constituido por la misma estructura de red, pero con asignación de Elo aleatorizada.

de los portales de Internet exhiben un Elo comparativamente menor al de los que juegan sobre tablero físico y muestran además una mayor variabilidad.

Asociada a cada conjunto de jugadores (sobre tablero físico y en Internet) construimos redes complejas en donde los nodos representan a los jugadores y las conexiones a la cantidad de partidas jugadas. Observamos que estas redes tienen distribuciones de grado heterogéneas, fuertemente sesgadas hacia la derecha. Esto implica que existe una gran variación en la cantidad de partidas –y por ende en el número de oponentes– que cada jugador enfrenta en su carrera ajedrecística. En el caso de los juegos en Internet (Portal), la distribución se ajusta bien con una ley de potencias con un exponente $\sim 1{,}5$, en gran parte del rango, con un leve decaimiento exponencial en la cola. Este decaimiento para grandes valores de k es fácil de entender ya que los jugadores están limitados en

la máxima cantidad de partidas debido a la duración de su carrera. En el caso de los jugadores sobre el tablero físico (OTB), la curva se aparta de una linea recta en un gráfico log-log, presentando una forma cóncava. El *plateau* para valores de k reducidos puede entenderse en términos de lo que mencionamos anteriormente (los jugadores que tienen poca actividad son por lo general aficionados de manera que sus partidas no suelen ser registradas).

Una de las principales diferencias entre las redes estudiadas se manifiesta en las correlaciones entre los grados de los nodos. Mientras que OTB es una red asortativa, como es el caso de muchas redes sociales [New02], Portal tiene una estructura disortativa. Esta diferencia puede interpretarse considerando, además del grado de los nodos, el nivel de juego de los jugadores (es decir, su Elo) como un atributo de los nodos. En OTB existe una correlación positiva entre estas dos variables, indicando que los jugadores mas activos son además los mas destacados. Dado que los jugadores enfrentan oponentes similares en fuerza, las asortatividad en términos del grado es natural. Por otra parte, en Portal no se observa una correlación entre el Elo de los jugadores y el grado de los nodos, de manera que es igualmente probable que un jugador enfrente oponentes con variados niveles de actividad. El estudio realizado sobre el fenómeno de club de ricos complementa la caracterización de ambas redes. Observamos que en las dos comunidades de ajedrez (OTB y Portal) los jugadores de élite tienden a formar grupos densamente conectados, lo cual se ve reflejado en el comportamiento creciente del coeficiente de club de ricos por Elo introducido en este capítulo.

En lo que respecta a la transitividad de las redes mostramos que ambas presentan un valor decreciente del coeficiente de clustering como función del grado de los nodos, aunque con algunas diferencias. En particular, vemos que el decrecimiento es bastante leve en OTB mientras que en Portal es bien pronunciado y puede ser aproximado por la relación $c(k) \sim k^{-1}$. Como se discute en Ravasz [Rav+02; RB03; Rav04], esta diferencia sugiere que la red Portal tiene una estructura jerárquica más pronunciada. Este resultado es consistente con el hecho de que la correlación entre el clustering y el grado no puede ser completamente removida cuando se realiza una aleatorización de la red Portal, indicando que la correlación es inherente a su topología. La presencia de una estructura jerárquica implica que los *hubs* (los nodos muy conectados) no forman una comunidad cerrada. Mas aún, esto concuerda con los resultados obtenidos en el análisis de club de ricos por grado, el cual muestra que en Portal este coeficiente decae a medida que aumenta el grado. El aislamiento de los *hubs* es posible pues estos están asociados a la clase de jugadores de nivel intermedio, la cual es muy popular en Portal. En los juegos de portales de Internet los jugadores que disputan una partida son apareados aleatoriamente del conjunto de jugadores registrados con Elo similar. Como hay pocos *hubs*, es poco probable que por este método los jugadores asociados a los mismos se enfrenten entre ellos, lo que explica que no formen comunidades cerradas.

La existencia de comunidades en redes OTB y Portal se contrastaron con experimentos realizados en modelos nulos apropiados, los cuales consisten en

generar versiones aleatorizadas de ambas redes, preservando sus distribuciones de grado. Como resultado de este análisis observamos que la modularidad de los modelos nulos es bastante menor que la obtenida en las redes originales, y que lo opuesto ocurre para del parámetro de mezcla. Es decir, la aleatorización de las redes borra la estructura de comunidades. El análisis de detección de comunidades revela además la existencia de correlaciones entre la estructura de comunidades de las redes y el Elo de los jugadores. Como regla general, el Elo de cada comunidad muestra una dispersión menor que la que se observa cuando se considera la red en su totalidad. Además, nuestro estudio realizado empleando el algoritmo propuesto por Newman y Clauset [NC16] (tomado al Elo medio de los jugadores como atributo) confirma que existe una correlación significativa entre la estructura de comunidades y el desempeño de los jugadores.

Capítulo 4

Ataques basados en centralidad sobre redes aleatorias

4.1. Resumen

En este capítulo analizaremos la robustez de redes de Erdős-Rényi ante cuatro tipos de ataques dirigidos guiados por medidas de centralidad. Las medidas de centralidad utilizadas son el grado y el betweenness y en los dos casos abordamos las versiones iniciales y recalculadas de los ataques. Realizamos este análisis comparativo empleando la teoría de los procesos de percolación, mediante la cual determinamos exponentes críticos analizando efectos de escala para tamaños finitos. Los exponentes de los ataques iniciales por grado y betweenness y el ataque recalculado por grado son consistentes con la percolación aleatoria, mientras que el ataque recalculado por betweenness se aparta de este comportamiento, presentando una transición abrupta con exponentes críticos no triviales, de manera semejante a los procesos de percolación explosiva.

4.2. Robustez de las redes con distribución de grado homogénea

Gran parte de la literatura de redes complejas estudia el comportamiento de las redes libres de escala. Esto es así porque este tipo de redes es frecuentemente observado en sistemas de diversas disciplinas. Específicamente, ha habido importantes trabajos en relación a la robustez de estas redes frente a fallas aleatorias y frente a ataques dirigidos [Hol+02; Cal+00; AJB00; Coh+01]. La principal conclusión que se extrae de estos trabajos es que las redes heterogéneas son robustas cuando se trata de fallas aleatorias, pero particularmente frágiles cuando los nodos son atacados según su conectividad. Este resultado no es difícil de entender. Una red heterogénea posee una gran cantidad de nodos periféricos, de grado bajo, cuyo rol en la conectividad del sistema es marginal. En un escenario en donde cada nodo tenga la misma probabilidad

de fallar, observaremos que gran parte de las fallas ocurren sobre estos nodos, sin comprometer de manera significativa la funcionalidad del sistema. Por otro lado, este tipo de redes presenta unos pocos nodos extremadamente conectados, los *hubs*, cuya remoción desencadena el colapso del sistema.

En contraposición, en redes con distribución de grado homogénea, tales como las de Erdös-Rényi (ER), las características estructurales de los nodos normalmente no difieren significativamente, por lo que es esperable que sean robustas frente a ataques dirigidos. En particular, esto es cierto para ataques basados en grado [AJB00]. Existen, sin embargo, algunos ejemplos de redes con distribución de grado homogénea para las cuales ciertos ataques dirigidos resultan muy efectivos. Tomemos, a modo de ejemplo, dos redes tecnológicas frecuentemente estudiadas: la red de distribución eléctrica de Estados Unidos (US-powergrid) [WS98] y la red de rutas interconectadas de Europa (Euroroad) [ŠB11]. Como podemos ver en la Figura 4.1(a-b), estas dos redes tienen una distribución de grado exponencial, sin la presencia de hubs[1], por lo que podemos incluirlas dentro de la categoría de redes homogéneas. La robustez frente a fallas y ataques se puede inferir de los paneles (c-d) de la figura. En estos paneles mostramos la evolución del tamaño relativo de la componente gigante S_1 como función de la fracción f de nodos removidos (ver Sección 2.4.2). Cada curva corresponde a un ataque diferente: betweenness recalculado (RB), grado recalculado (RD), betweenness inicial (IB), grado inicial (ID), y remoción aleatoria (Rnd), equivalente a percolación estándar. En ambos casos observamos que las fallas aleatorias generan el menor daño en las redes. Los ataques dirigidos, si bien son todos más efectivos, se comportan de manera diferente entre sí. En particular, el ataque por betweenness recalculado resulta extremadamente efectivo, destruyendo ambas redes con menos del 2 % de los nodos atacados. Además, la ruptura generada por este ataque parece ser más abrupta que la de los otros ataques dirigidos.

El resultado anterior sugiere que las redes homogéneas no son universalmente robustas ante ataques dirigidos. Dicho de otro modo, los sistemas pueden presentar vulnerabilidades importantes incluso en ausencia de *hubs*, y estas pueden ser explotadas por alguna estrategia de ataque adecuada. En lo que sigue del capítulo intentaremos formalizar este fenómeno. Para ello, estudiaremos en detalle las estrategias de ataque mencionadas sobre el modelo más simple de redes homogéneas: las redes de Erdös-Rényi.

[1] Esta característica tiene que ver con el hecho de que ambas redes están embebidas en el plano, lo cual impone restricciones a la estructura de conexiones, como veremos en el próximo capítulo.

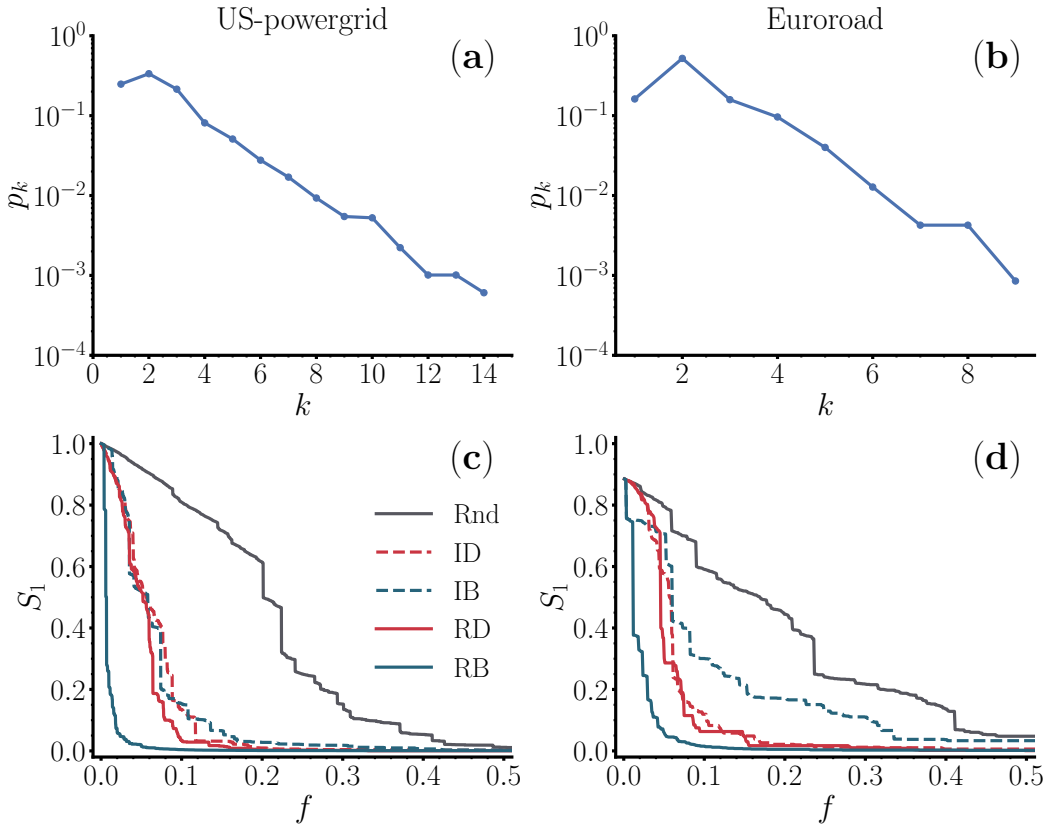

Figura 4.1: Robustez frente a diferentes estrategias de ataque sobre redes empíricas con distribución de grado homogénea. Los paneles de la izquierda corresponden a la red de distribución de energía eléctrica de Estados Unidos y los de la derecha a la red de rutas terrestres de Europa. **(a-b)** Distribución de grado. **(c-d)** Evolución de la componente gigante como función de la fracción de nodos removidos.

4.3. Análisis cualitativo de las transiciones de percolación

En la Figura 4.2 mostramos un análisis similar al de la Figura 4.1, pero para una red ER con grado medio $\langle k \rangle = 5$. Cada curva, en este caso, representa un promedio tomado sobre 10^3 redes independientes y cada panel corresponde a un tamaño de red diferente. Cuando el tamaño de la red es pequeño (panel (a)), podemos ver que ID resulta ligeramente mejor que IB, en el sentido de que, para cada valor de f, la red está consistentemente más fragmentada cuando los nodos de alto grado son removidos. Tal como es mencionado por Iyer, et al. en [Iye+13], la situación se revierte cuando cuando la lista de nodos es recalculada luego de cada paso, de manera que RB resulta más efectivo que RD. Cuando atacamos una red más grande con las mismas características (panel (b)), las

transiciones se definen de mejor manera. Además, todas parecen consistentes con transiciones de fase continuas, a excepción de RB, cuya continuidad no resulta clara a simple vista. Este último ataque produce un colapso abrupto de la red a $f \approx 0{,}3$, con una pendiente muy pronunciada. Sin embargo, para valores menores de f este ataque resulta poco efectivo (ver recuadro), siendo apenas mejor que la remoción aleatoria. De hecho, en el inicio del proceso ambos ataques (RB y Rnd) no producen una fragmentación significativa de la red, tal como se puede ver al comparar con la curva negra, correspondiente al ataque menos efectivo[2].

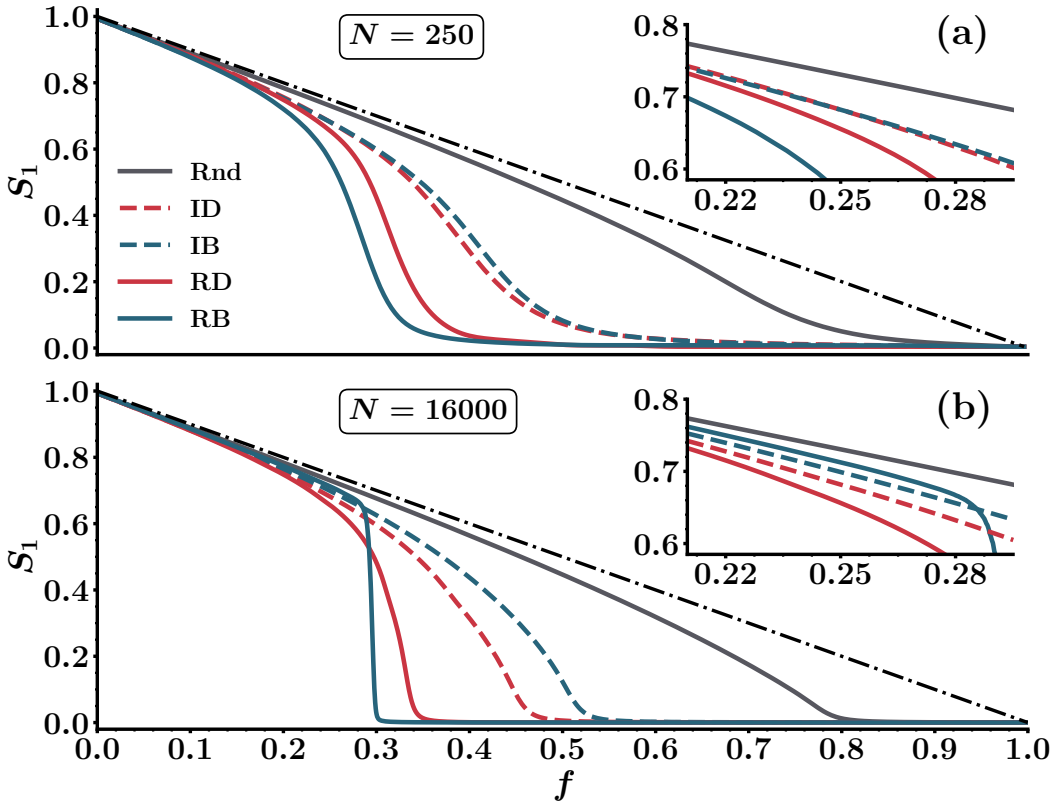

Figura 4.2: Tamaño relativo de la componente gigante como función de la fracción de nodos removidos, promediado sobre 10^3 realizaciones independientes, para dos redes de tipo ER con $\langle k \rangle = 5$ y tamaño N. A modo de referencia, la línea punteada corresponde a la remoción de nodos utilizando el ataque menos efectivo posible (ver texto principal). El recuadro muestra en detalle el comportamiento de los ataques justo antes del colapso producido por RB. (a) $N = 250$, (b) $N = 16000$.

[2]Con "ataque menos efectivo", nos referimos a remover, a cada paso, un nodo que no genere divisiones en la red. Esto siempre es posible y el valor del parámetro de orden para este ataque es $S_1(f) = 1 - f$.

4.4. Determinación del umbral de percolación

Como mencionamos en la Sección 2.4.2, el umbral de percolación es una de las características distintivas en una transición de fase de percolación. A diferencia de los exponentes críticos, que toman el mismo valor para los múltiples modelos que integran una misma clase de universalidad, el valor del umbral en general depende de los detalles del sistema. Al cambiar alguna característica de la red o del mecanismo de percolación, es esperable que el umbral de percolación se vea modificado. Debido a este carácter no universal, se han desarrollado numerosos métodos, tanto analíticos como numéricos, para determinar con precisión este parámetro. Entre los métodos analíticos se encuentran los métodos descriptos en el Apéndice C, basados en el cálculo de funciones generatrices. En esta sección, discutiremos algunos métodos basados en simulaciones numéricas.

La mayoría de los métodos desarrollados para determinar el umbral de percolación basados en simulaciones requiere simular sistemas de gran tamaño (ver Sección 2.4.3). En percolación aleatoria esto no suele ser un problema, ya que existen algoritmos muy eficientes para realizar las simulaciones [NZ00; NZ01]. Cuando se trabaja con algoritmos de alta complejidad computacional, como es el caso del cálculo de betweenness, ya no es posible simular sistemas grandes. Al usar los métodos tradicionales con sistemas de tamaño pequeño, los errores estadísticos pueden ser significativos, por lo que el cálculo del umbral de percolación se vuelve menos preciso. Algunos métodos han sido desarrollado específicamente para sistemas pequeños, como por ejemplo el método propuesto en [Bas+14]. Desarrollaremos a continuación uno de ellos, introducido en el artículo [ABP20] asociado a esta tesis, con el cual obtuvimos los resultados más precisos.

Consideremos la hipótesis de escala (2.42), que describe los tamaños de las componentes $S_i(f, N)$ para una dada fracción f de nodos removidos en un sistema de tamaño N[3] y definamos $Q(f, N)$ como el cociente entre las dos componentes más grandes. Es decir,

$$Q(f, N) = \frac{S_1(f, N)}{S_2(f, N)}. \tag{4.1}$$

La magnitud Q depende, en general, tanto del parámetro de control f como del tamaño N. Sin embargo, si tenemos en cuenta la ecuación (2.42), podemos verificar que, en el punto crítico, $Q(f_c, N) \sim \tilde{S}_1(0)/\tilde{S}_2(0)$. En otras palabras, el valor de Q en el umbral de percolación es el mismo para todo tamaño. Como consecuencia de esto, las curvas $Q(f, N)$ para diferentes tamaños deben intersectarse justo en el punto crítico. El método, al que daremos el nombre de *método de cruce*, consiste entonces en estimar numéricamente la posición de esta intersección. Para ello, utilizamos promedios tomados sobre $N_s = 2 \times 10^4$ redes independientes para los ataques ID, RD e IB, y sobre $N_s = 5 \times 10^3$

[3] Si bien la Ecuación (2.42) está formulada en términos del tamaño lineal L y no de N, el método aquí descripto es igualmente válido en ambos casos

redes independientes para RB de distintos tamaños y calculamos los cocientes $Q(f, N)$. Luego determinamos la posición de la intersección para cada par de tamaños. Los valores de f_c que reportamos corresponden a los valores medios de las intersecciones, con la desviación estándar como medida de la incerteza asociada. Los resultados del método, aplicado a las cuatro estrategias de ataque descriptas anteriormente, están ilustrados en la Figura 4.3 y resumidos en la Tabla 4.1. Tal como se puede ver en la figura, la exactitud del método depende de la naturaleza del ataque. Para los dos ataques basados en grado, la intersección está muy bien definida, por lo que la desviación estándar es relativamente baja. Para estos casos obtuvimos una estimaciones precisas de los umbrales de percolación, a saber $f_c^{\mathrm{ID}} = 0{,}4652(7)$ y $f_c^{\mathrm{RD}} = 0{,}3401(2)$. En el ataque por betweenness inicial, en cambio, la precisión es menor, y el valor obtenido fue $f_c^{\mathrm{IB}} = 0{,}558(1)$. Por último, se puede ver que el método funciona muy bien para evaluar el ataque con betweenness recalculado, incluso cuando el número de simulaciones y los tamaños alcanzados son menores (debido a la complejidad computacional del algoritmo), obteniéndose el valor $f_c^{\mathrm{RB}} = 0{,}2984(2)$.

Desde la perspectiva de desmantelamiento de redes, las versiones recalculadas de los ataques resultan más efectivas que sus respectivos ataques iniciales, ya que tienen menores umbrales de percolación[4]. En particular, la versión inicial del ataque por betweenness evidencia ser una estrategia de ataque muy pobre, superando apenas la remoción aleatoria de nodos. Por otra parte, la correspondiente versión recalculada de este ataque es la más eficiente, siendo incluso comparable con las mejores estrategias de ataque conocidas [Bra+16; MM15].

4.5. Exponentes críticos asociados a los ataques

Una vez estimado el umbral de percolación, procederemos a determinar los exponentes críticos asociados a cada estrategia de ataque. Comenzaremos, para ello, estudiando el comportamiento de algunos de los observables de la transición para distintos tamaños, de manera de explotar las relaciones de escala introducidas en la Sección 2.4.3.

Los cuatro paneles de la Figura 4.4 muestran el parámetro de orden S_1 como función de la fracción de nodos removidos. Cada panel corresponde a un ataque diferente y cada curva, a un tamaño de red diferente. Se puede observar que las transiciones se vuelven más abruptas en la medida en que el tamaño crece, particularmente en el caso del ataque RB. Las curvas dentro de cada panel se pueden colapsar en una curva maestra para ambos lados de la transición, como mostramos en los recuadros de cada panel, haciendo uso de la

[4]Este resultado, si bien parece una obviedad, no se cumple para todo tipo de redes y ataques. En particular, Holme, et al. discuten en [Hol+02] un ejemplo para el cual ID resulta más efectivo que RD.

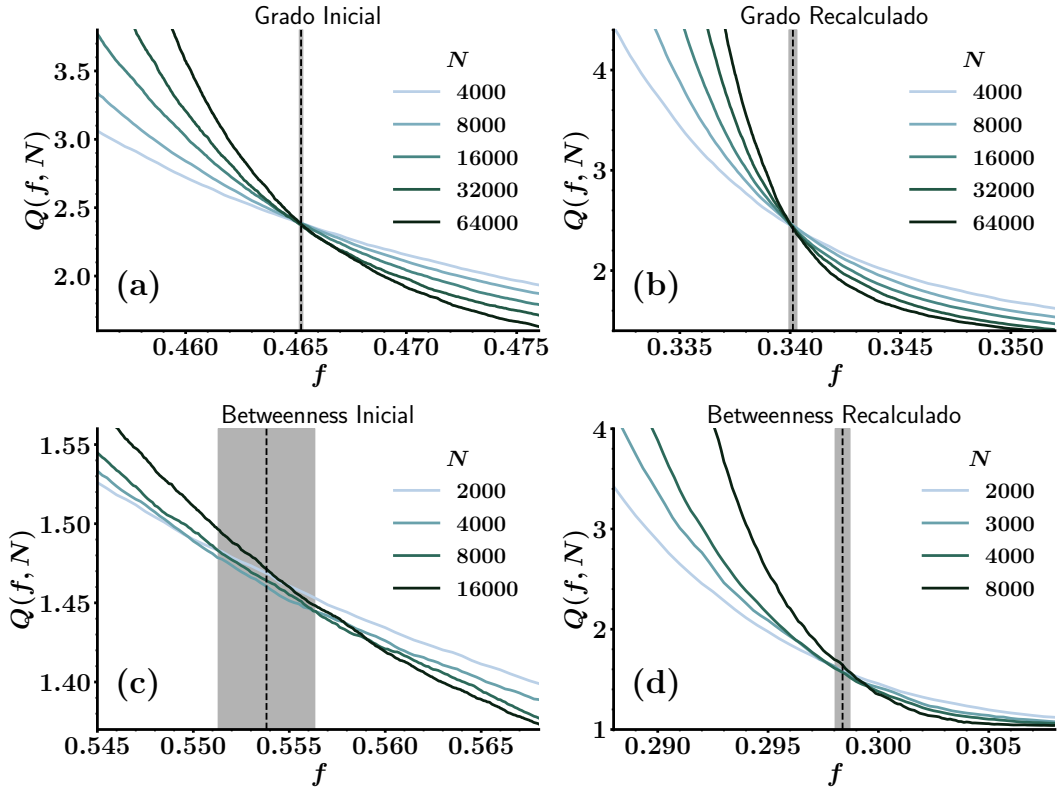

Figura 4.3: Determinación del umbral de percolación f_c empleando el método de cruce para las cuatro estrategias de ataque consideradas. Cada curva representa un promedio sobre N_s realizaciones independientes. La línea vertical corresponde a la media de las posiciones de los cruces y la región sombreada a la desviación estándar. Los valores obtenidos están indicados en la Tabla 4.1. **(a)** Grado inicial (ID), $N_s = 2 \times 10^4$, **(b)** Grado recalculado (RD), $N_s = 2 \times 10^4$, **(c)** Betweenness inicial (IB), $N_s = 2 \times 10^4$, **(d)** Betweenness recacalculado (RB), $N_s = 5 \times 10^3$.

hipótesis de escala (2.42)[5]. Los exponentes $\beta/\bar{\nu}$ y $\bar{\nu}$ empleados para colapsar las curvas correspondientes a cada ataque, los cuales fueron estimados empleando los métodos que describiremos a continuación, están expresados en la Tabla 4.1. Para los cuatro ataques se puede observar que las curvas colapsan muy bien en curvas maestras en la cercanía del punto crítico, lo cual confirma la validez de las relaciones de escala.

En las Figuras 4.5 y 4.6 se muestran el tamaño de la segunda componente $S_2 N$ y la susceptibilidad $\langle s \rangle$. Estas dos cantidades exhiben un pico en la vecindad de la transición, el cual aumenta en magnitud a medida que aumenta el tamaño del sistema. De la misma manera que sucede para el parámetro de

[5]Estrictamente, utilizamos una versión equivalente a (2.42) expresada en términos del tamaño N, como en la Ecuación (2.41)

	f_c	$\beta/\bar{\nu}$	$\gamma/\bar{\nu}$	$\bar{\nu}$	τ
Rnd	0,8	1/3	1/3	3	2,5
ID	0,4652(7)	0,320(4)	0,354(5)	2,72(5)	2,50(2) [2,48(2)]
RD	0,3401(2)	0,307(3)	0,377(7)	2,59(7)	2,43(3) [2,45(2)]
IB	0,558(1)	0,340(3)	0,334(5)	2,8(2)	2,52(2) [2,50(2)]
RB	0,2984(2)	0,10(1)	0,89(2)	1,50(5)	— [2,1(2)]

Cuadro 4.1: Estimación numérica para los umbrales de percolación y exponentes críticos correspondientes a diferentes estrategias de ataques efectuadas sobre una red de tipo ER con $\langle k \rangle = 5$. Los valores de τ entre corchetes fueron obtenidos utilizando la Ecuación (2.32).

orden, las curvas para distintos tamaños pueden colapsarse en una única curva maestra utilizando las ecuaciones (2.42) y (2.43). Los recuadros correspondientes confirman la validez de las relaciones de escala.

Nos enfocamos ahora en la estimación de los exponentes críticos. Evaluando la Ecuación (2.43) en $f = f_c$, tenemos que $\langle s \rangle (f_c, N) \sim N^{\gamma/\bar{\nu}}$, por lo que si graficamos en escala log-log $\langle s \rangle$ vs N en el punto de percolación, deberíamos ver una recta con pendiente $\gamma/\bar{\nu}$. Luego, este cociente entre exponentes puede obtenerse mediante un ajuste lineal de la recta. La desventaja principal de este método es que requiere conocer el umbral de percolación de antemano. Como nosotros contamos sólo con una estimación de f_c, este método propagaría la incerteza asociada a la estimación. Para evitar esto, en lugar de calcular el tamaño medio de componentes en el punto crítico, calculamos el valor del pico para cada tamaño, teniendo en cuenta que para tamaños del sistema suficientemente grandes el escaleo de los picos resulta idéntico al escaleo en el punto crítico [RCR18].

En la Figura 4.7 mostramos los escaleos correspondientes a cada una de las estrategias de ataque implementadas. En todos los casos, la relación lineal es claramente visible. Los cocientes entre exponentes críticos estimados se muestran en la figura y están incluidos en la Tabla 4.1.

Empleando la misma metodología, utilizamos el pico en $S_2 N$ para obtener el cociente $\beta/\bar{\nu}$. De acuerdo con la ecuación (2.42), el pico escala como $S_2 N \sim N^{1-\beta/\bar{\nu}}$ cerca del umbral de percolación. En la Figura 4.7 se incluyen los valores de los exponentes calculados, y en la Tabla 4.1, los correspondientes valores de $\beta/\bar{\nu}$.

Como mencionamos en la Sección 2.4.2, en percolación aleatoria sólo dos de los exponentes críticos son independientes, mientras que los demás pueden ser obtenidos a partir de las relaciones de escala e hiperescala. Al emplear estrategias de ataque, las hipótesis sobre las cuales se deducen estas relaciones no necesariamente se cumplen, por lo que su validez no puede ser asegurada de antemano. Tomando los cocientes entre exponentes calculados hasta ahora podemos analizar la validez de la relación de hiperescala (2.33). Los valores

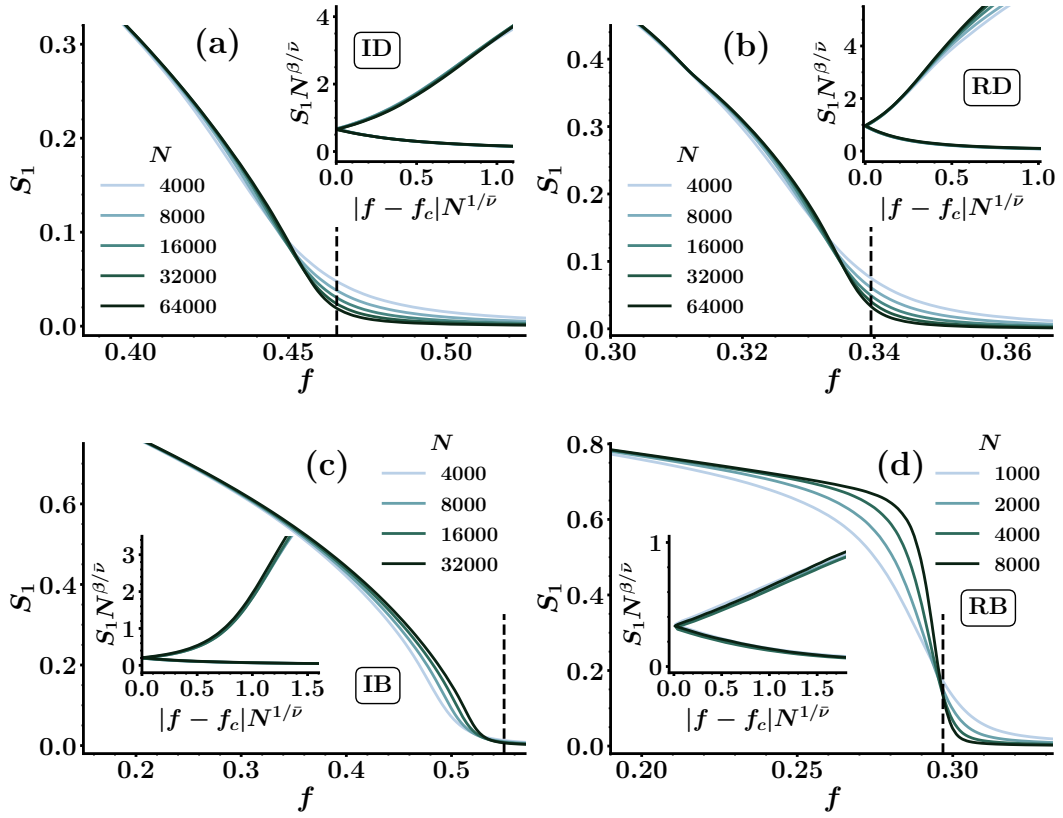

Figura 4.4: Parámetro de orden S_1 como función de la fracción f de nodos removidos, en la cercanía del umbral de percolación. Cada panel corresponde a una de las cuatro estrategias de ataque estudiadas. Las líneas punteadas verticales indican el valor del umbral de percolación, calculado mediante el método de cruce (ver texto principal). Los recuadros muestran el colapso de las curvas utilizando la hipótesis de escala (2.42). Los valores de los exponentes críticos empleados para realizar el colapso están resumidos en la Tabla 4.1. **(a)** ID, **(b)** RD, **(c)** IB, **(d)** RB.

obtenidos al reemplazar los términos del lado izquierdo de la ecuación por los valores calculados fueron 0,99(2) para ID, 0,99(2) para RD, 1,01(2) para IB y 1,09(4) para RB. De los cuatro casos, tres satisfacen la igualdad, mientras que sólo uno (la versión recalculada de betweenness) presenta un valor que se desvía ligeramente de 1. Debido a que los tamaños simulados para este ataque son pequeños, este resultado debe ser tomado con cautela, por lo que no podemos afirmar que esta diferencia sea significativa.

Con los métodos empleados hasta ahora, sólo hemos podido obtener relaciones entre exponentes ($\beta/\bar{\nu}$ y $\gamma/\bar{\nu}$). Sin embargo, es posible obtener el exponente $\bar{\nu}$ a partir de la relación de escala (2.42) y así determinar los valores de β y de

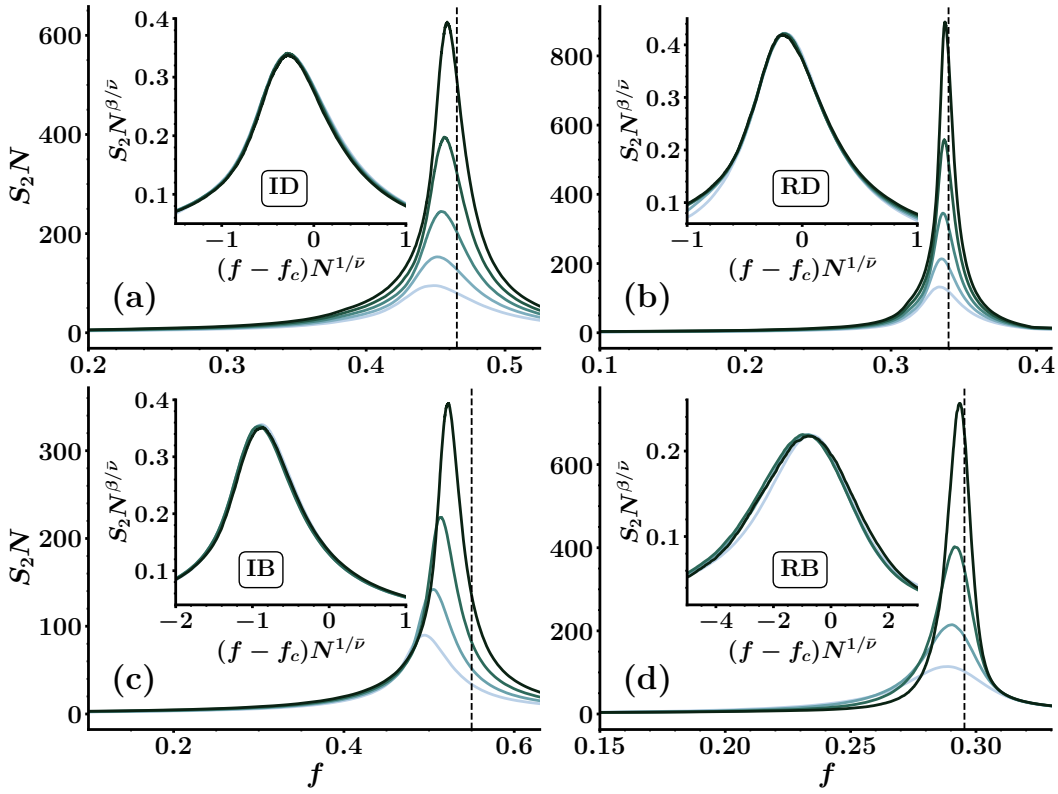

Figura 4.5: Tamaño de la segunda componente S_2N como función de la fracción f de nodos removidos. Como se puede ver, este observable presenta un pico en la cercanía del punto crítico (línea vertical punteada). El código de colores empleado es el mismo que el de la Figura 4.4. Los recuadros muestran el colapso de las curvas utilizando el ansatz de la Ecuación (2.42), con los parámetros reportados en la Tabla 4.1. **(a)** ID, **(b)** RD, **(c)** IB, **(d)** RB.

γ. Definamos, para ello, la función

$$G(f, N) = -\frac{\partial \log S_1(f, N)}{\partial f}. \tag{4.2}$$

De acuerdo con (2.42), en torno al punto crítico $G(f, N)$ tomará la forma

$$\begin{aligned} G(f, N) &\sim -\frac{\partial \ \log \tilde{S}_1 \left[(f - f_c)N^{1/\bar{\nu}}\right]}{\partial f} \\ &\sim N^{1/\bar{\nu}} \tilde{G} \left[(f - f_c)N^{1/\bar{\nu}}\right], \end{aligned} \tag{4.3}$$

donde $\tilde{G}(x) = -\tilde{S}_1'(x)/\tilde{S}_1(x)$. Luego, la función G cumple con una relación de escala similar a la del parámetro de orden o la susceptibilidad, sólo que en este caso el exponente crítico asociado es igual a 1. Además, como el parámetro de orden tiene un punto de inflexión en torno al umbral de percolación, entonces G

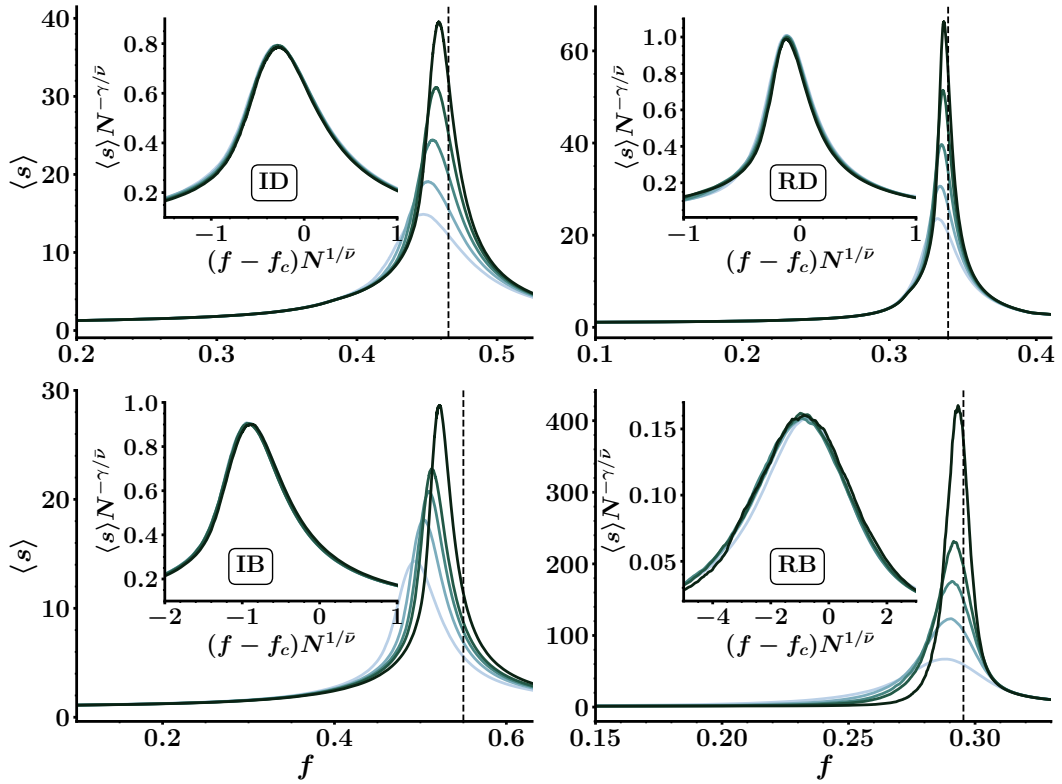

Figura 4.6: Susceptibilidad $\langle s \rangle$ como función de la fracción f de nodos removidos para las cuatro estrategias de ataque estudiadas. De la misma manera que sucede para la segunda componente, la susceptibilidad presenta un pico en la vecindad de la transición. Los recuadros muestran el colapso de las curvas empleando el ansatz dado por la Ecuación (2.43), con los parámetros reportados en la Tabla 4.1. **(a)** ID, **(b)** RD, **(c)** IB, **(d)** RB.

debe tener un máximo en ese punto. Al igual que sucede con la susceptibilidad y con la segunda componente más grande, es esperable que los picos de $G(f, N)$ escalen como una ley de potencias, en este caso con exponente asociado $1/\bar{\nu}$. Por lo tanto, podemos emplear un análisis similar al realizado anteriormente, graficando en escala log-log el pico de G en función de N, donde deberíamos ver una relación lineal. Sin embargo, debemos hacer una salvedad en relación a esta estrategia. En general, tomar la derivada numérica de una señal ruidosa (como lo son los resultados de las simulaciones) tiende a amplificar el ruido existente. Nuestro caso no es la excepción, como puede apreciarse en los paneles izquierdos de la Figura 4.8. Las curvas grises en cada panel corresponden a la función $G(f, N)$ estimando la derivada mediante diferencias finitas de cinco puntos (la cantidad de realizaciones es la misma que para las figuras anteriores). Podemos ver que el ruido es amplificado y que además incrementa con el tamaño de la red. Para solucionar este problema, utilizamos un método de regularización, descripto

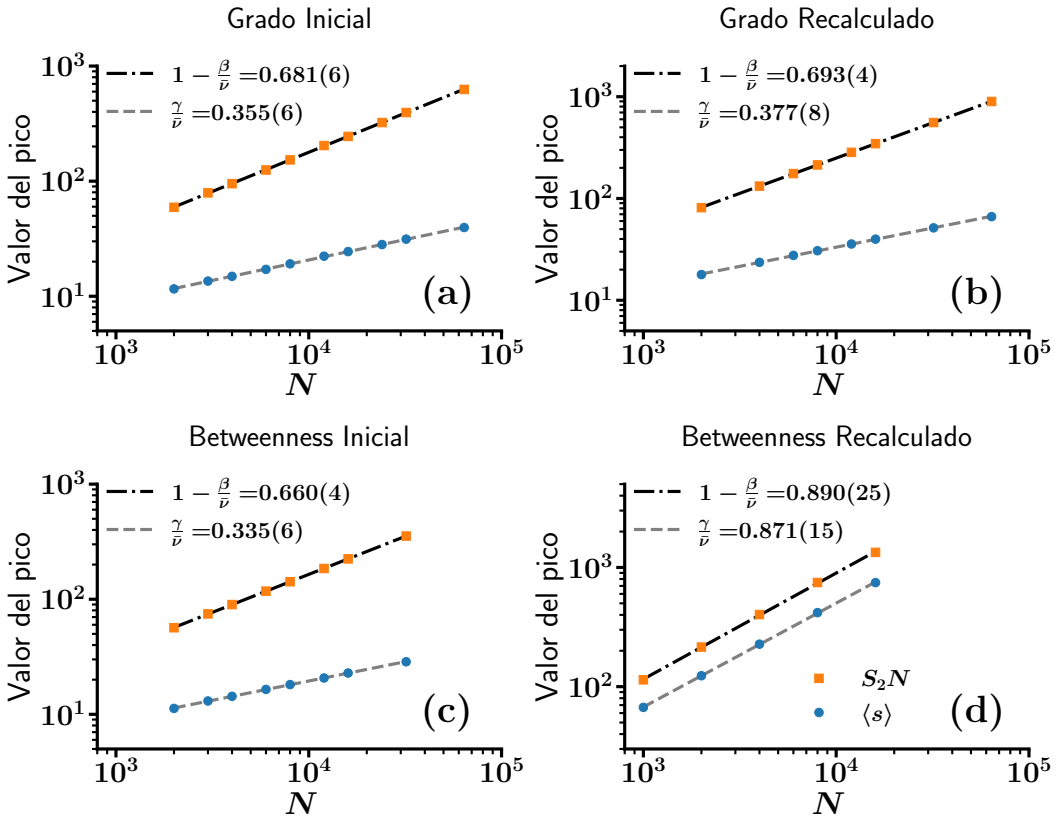

Figura 4.7: Escaleo de los picos de S_2N y $\langle s \rangle$ para las cuatro estrategias de ataque empleadas. Los símbolos representan los valores de los picos, luego de promediar sobre N_s simulaciones. Las líneas punteadas corresponden al ajuste lineal por cuadrados mínimos realizado sobre la escala log-log. **(a)** ID, **(b)** RD, **(c)** IB, **(d)** RB.

en [Cha11], con el cual pueden obtenerse las curvas suavizadas coloreadas de la figura. Los paneles derechos de la Figura 4.8 muestran el escaleo de los picos, calculados a partir de las derivadas regularizadas. Como podemos ver en las regresiones lineales, la hipótesis de escala (4.3) se satisface. Los valores obtenidos para el exponente de la longitud de correlación fueron $\bar{\nu}^{(\mathrm{ID})} = 2{,}72(5)$, $\bar{\nu}^{(\mathrm{RD})} = 2{,}59(7)$, $\bar{\nu}^{(\mathrm{IB})} = 2{,}8(2)$ y $\bar{\nu}^{(\mathrm{RB})} = 1{,}50(5)$.

Para tener una referencia en relación a estos resultados, la longitud de correlación en el proceso de percolación aleatoria en redes de Erdös-Rényi diverge con exponente $\nu = 1/2$, mientras que la dimensión efectiva asociada es $d = 6$ [SA18]. Por lo tanto, en percolación aleatoria tenemos $\bar{\nu} = 3$. Nuevamente, vemos que tres de los ataques estudiados (ID, RD e IB) se asemejan a este valor, mientras que RB difiere significativamente.

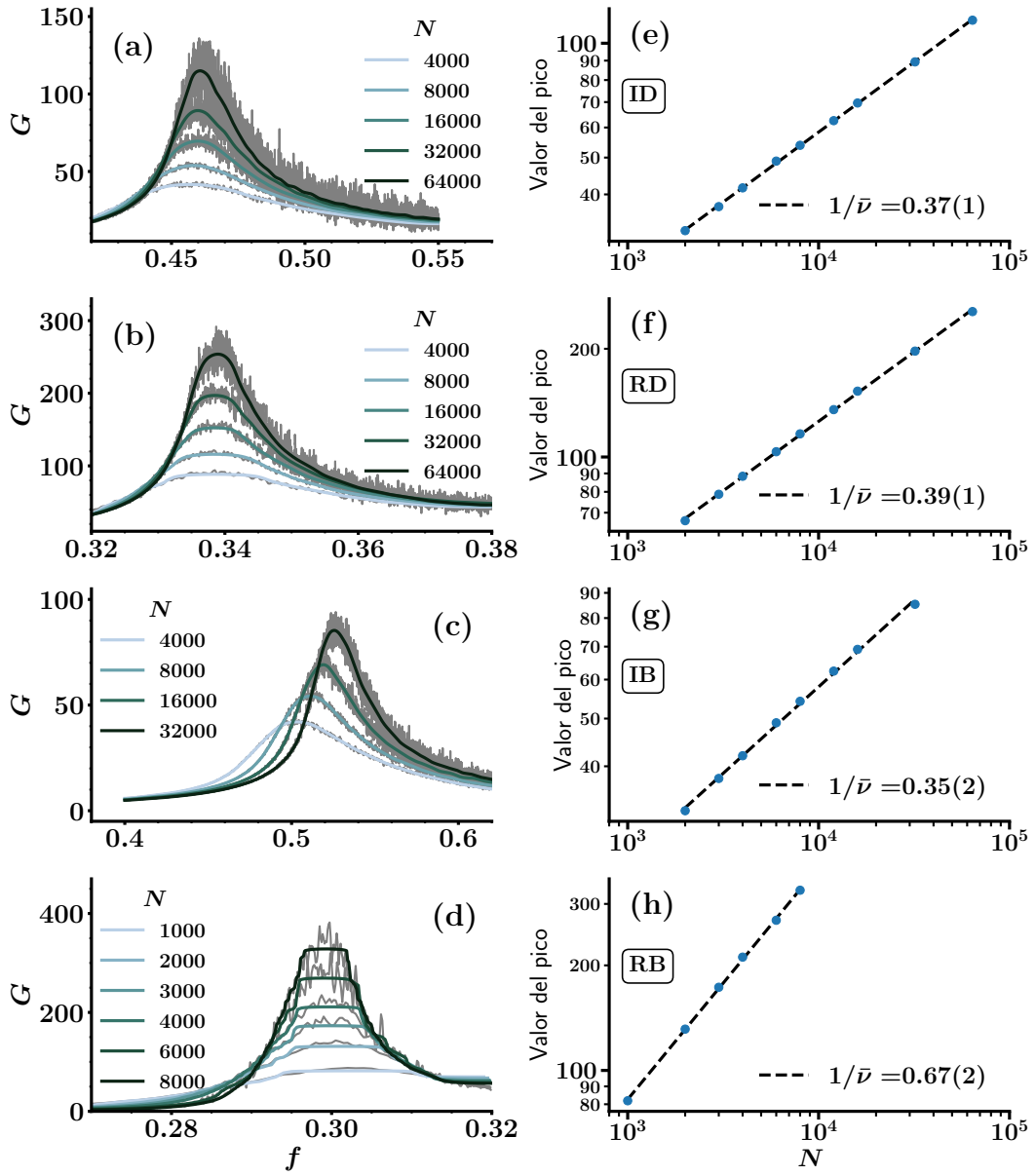

Figura 4.8: **(a-d)** Derivada del logaritmo del parámetro de orden en función de la fracción de nodos removidos. las curvas grises corresponden a la derivada numérica de cinco puntos, y las curvas coloreadas a las derivadas calculadas utilizando el método de regularización descripto en [Cha11]. Ambos métodos fueron aplicados sobre promedios realizados sobre $N_s = 2 \times 10^4$ simulaciones (ID, RD e IB), y $N_s = 5 \times 10^3$ simulaciones para RB. **(e-h)** Escaleo de los picos de las curvas de la izquierda. Las líneas punteadas corresponden a ajustes lineales por cuadrados mínimos. Los valores obtenidos para el exponente crítico $1/\bar{\nu}$ están resumidos en la Tabla 4.1.

4.6. Determinación semi-analítica de f_c en los ataques por grado

En la Sección 2.2.2 introdujimos el modelo de configuración como herramienta matemática que permite calcular, a nivel estadístico, muchas de las propiedades de ensambles de redes con una dada distribución de grado. La principal hipótesis en la que se basa este modelo es la de ausencia de correlaciones de grado. Es decir, las predicciones del modelo sólo son exactas si las redes del ensamble están perfectamente decorrelacionadas. Afortunadamente, las redes de Erdös-Rényi cumplen con esta hipótesis, ya que la probabilidad de conexión entre un dado par de nodos es independiente de las otras conexiones existentes.

El fenómeno de percolación en redes decorrelacionadas puede ser abordado de manera analítica utilizando el modelo de configuración [New18]. Si bien existen distintos enfoques teóricos, uno de los más simples consiste en utilizar la teoría de *funciones generatrices* [Wil05]. Para el caso de percolación aleatoria es posible determinar tanto el umbral de percolación f_c como los exponentes críticos [New18, Capítulo 15]. Cuando se trata de ataques dirigidos, la posibilidad de llegar a algún resultado de manera analítica o semi-analítica depende de las características del ataque en cuestión. Como veremos a continuación, es posible obtener algunos resultados para los ataques por grado, tanto en la estrategia inicial como en la recalculada.

La versión inicial del ataque por grado fue abordada por Callaway, et al. en [Cal+00], extendiendo las ideas de percolación aleatoria y empleando la teoría de funciones generatrices. Los aspectos más relevantes de este trabajo están resumidos en la Sección C.2 del Apéndice C. En particular, el umbral de percolación puede obtenerse utilizando la expresión (C.17) para resolver numéricamente la igualdad (C.15). Utilizando métodos iterativos, resolvimos la Ecuación (C.15) para el caso de redes de Erdös-Rényi con $\langle k \rangle = 5$. La solución obtenida fue $f_c^{\mathrm{ID}} = 0{,}46532(5)$, la cual coincide con el valor hallado mediante simulaciones numéricas utilizando el método del cruce.

La versión recalculada del ataque fue abordada de manera semi-analítica por Kim, et al. [KKG20] de manera paralela al desarrollo de esta tesis, y está resumida en la Sección C.3 del Apéndice C. En su trabajo, los autores encontraron una forma de expresar la evolución de la distribución de grado durante el transcurso del ataque mediante un conjunto de ecuaciones acopladas. Resolviendo numéricamente estas ecuaciones y empleando el criterio para la existencia de la componente gigante desarrollado en C.1, es posible estimar el umbral de percolación como $f_c^{\mathrm{RD}} = 0{,}34013(5)$. Al igual que en el ataque inicial, el valor obtenido coincide con el calculado utilizando el método del cruce.

Además de los umbrales de percolación, los métodos de funciones generatrices permiten obtener el valor del parámetro de orden S_1 a lo largo de todo el ataque. En la Figura 4.9 podemos ver que las curvas obtenidas resolviendo numéricamente las ecuaciones (C.13) para el ataque por grado inicial, y (C.18) para el ataque por grado recalculado coinciden con los resultados obtenidos en las

simulaciones numéricas (en este caso, mostramos la simulación correspondiente a una red de tamaño $N = 10^6$, para la cual los efectos de tamaño finito no se aprecian significativamente a simple vista).

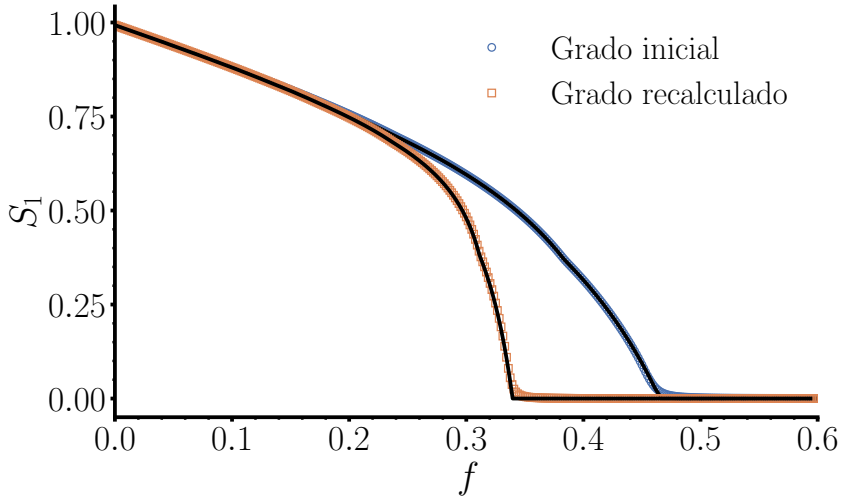

Figura 4.9: Estimación semi-analítica del tamaño de la componente gigante para los ataques por grado. Los círculos celestes muestran el valor de S_1 para una simulación del ataque por grado inicial en una red de $N = 10^6$ nodos. La línea negra superpuesta muestra el valor calculado utilizando la estimación basada en funciones generatrices (Ecuación (C.13). Se puede observar que el acuerdo entre ambos resultados es excelente. De la misma manera, los cuadrados naranja muestran la simulación del ataque por grado recalculado para la misma red. La linea negra superpuesta representa en este caso la estimación realizada utilizando el método de Kim, et al. (Ecuaciones (C.18)).

4.7. Distribución de tamaños de componentes

Como se explicó en la Sección 2.4.2, en el punto crítico las transiciones de percolación de segundo orden exhiben una distribución de tamaños de componentes finitas de tipo ley de potencias, descripta por la ecuación (2.31). En la Figura 4.10 mostramos que esto es así para los dos ataques basados en grado y para el ataque de betweenness inicial. Los exponentes de las respectivas leyes de potencia, los cuales fueron obtenidos directamente de n_s mediante un ajuste lineal por mínimos cuadrados en escala log-log, concuerdan dentro de la incerteza con el valor correspondiente a percolación aleatoria, $\tau = 2{,}5$ (ver Tabla 4.1). El caso de betweenness recalculado merece especial atención, ya que se desvía del comportamiento típico. Aunque se puede observar una caída de tipo ley de potencias para tamaños de componentes pequeños, la distribución presenta un *plateau* para tamaños intermedios. En conjunto con la caída abrupta del parámetro de orden, este comportamiento podría ser un

indicador de que la transición es de primer orden. Sin embargo, es importante resaltar que este tipo de efectos también se ve en algunas transiciones continuas, particularmente en el contexto de percolación explosiva [DN15; CZD12].

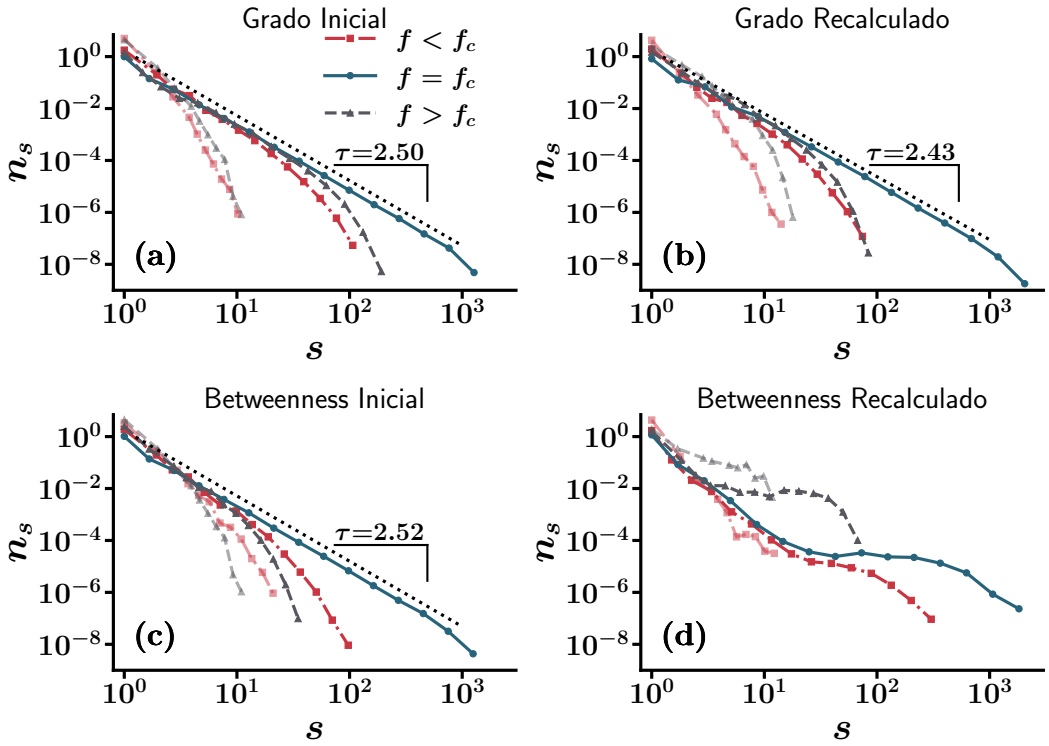

Figura 4.10: Distribución normalizada de tamaños de componentes $n(s)$ para cada una de las estrategias de ataque consideradas. Las líneas punteadas rojas corresponden a la región subcrítica $f < f_c$, las curvas punteadas grises a la región supercrítica $f > f_c$ y la línea sólida azul al punto crítico $f = f_c$. Para ID, RD e IB, los gráficos son consistentes con la ecuación (2.31), mostrando una distribución de tipo ley de potencias en el punto crítico (con una caída en la cola de la distribución debido a efectos de tamaño finito), y una caída exponencial para otros valores del parámetro de orden. Las líneas punteadas negras corresponden a ajustes lineales por mínimos cuadrados realizados sobre las curvas en el umbral de percolación. Los exponentes τ asociados se muestran en la imagen, como así también en la Tabla 4.1. Las distribuciones fueron construidas a partir de $N_s = 10^3$ realizaciones. (a) ID, $N = 64000$, (b) RD, $N = 64000$, (c) IB, $N = 64000$, (d) RB, $N = 16000$.

A continuación, daremos un argumento en favor de que este *plateau* es en realidad un efecto transitorio, consecuencia de los tamaños finitos estudiados, y que debería desaparecer en sistemas más grandes. Utilizando un argumento heurístico similar al de [DN15] podemos estimar el tamaño de *crossover* N^*, para el cual el sistema es lo suficientemente grande como para que las realizaciones

individuales converjan al comportamiento asintótico. Consideremos para ello el mayor salto en el parámetro de orden luego de remover un nodo en una realización individual $\Delta^{(i)}$, definido por la Ecuación (2.45). La variación en el parámetro de control f en este paso simple es $\Delta f = 1/N$. Asumiendo que este salto ocurre a $f = f_c$ y usando el escaleo del parámetro de orden, podemos decir que $\Delta^{(i)} \sim (\Delta f)^{-\beta} = N^{\beta}$. Ahora definimos N^* como el tamaño del sistema para el cual el salto más grande en el parámetro de orden es alrededor de un 10 %. Entonces, $N^* \sim 10^{1/\beta}$. Para el ataque por RB, $\beta \sim 0{,}15$, por lo que $N^* \sim 10^6$. Dado que los resultados presentados en la Figura 4.10 corresponden a $N = 16000$, nuestro sistema se encuentra aún por debajo del tamaño de *crossover*, lo cual podría explicar la desviación con respecto a la ley de potencia esperada.

4.8. Discusión

La Tabla 4.1 resume los principales resultados de este capítulo, proveyendo una caracterización completa de las transiciones de percolación producidas por las cuatro estrategias de ataque estudiadas. Desde la perspectiva de la problemática de desmantelamiento de redes, la magnitud relevante es el umbral de percolación, el cual cuantifica la efectividad de cada ataque. Nuestros resultados no sólo confirman que las versiones recalculadas de los ataques resultan más efectivas que sus contrapartes iniciales (aspecto que ha sido observado con anterioridad en la literatura [Hol+02; Iye+13; Wan+18]) sino que nos permite cuantificar el grado de mejora que se obtiene al recalcular la centralidad de los nodos después de cada paso. Si comparamos los dos ataques basados en grado, la diferencia entre sus correspondientes umbrales de percolación es de $\sim 0{,}12$. Para el caso de los ataques basados en betweenness, la mejora es de $\sim 0{,}26$. Como podemos ver, la diferencia es mayor en este último caso, y la causa de esto puede ser atribuida a la naturaleza global de la centralidad de betweenness, la cual contrasta con la localidad de la centralidad de grado.

Desde el punto de vista de la teoría de fenómenos críticos, los observables más importantes son los exponentes críticos, ya que determinan la clase de universalidad de la transición. Nuestros resultados muestran que los dos ataques basados en grado, junto con el ataque de betweenness inicial, pertenecen a la misma clase de universalidad que percolación estándar. Por otra parte, el ataque basado en betweenness recalculado se diferencia cualitativamente de los otros ataques. Los exponentes críticos difieren de los de valores de campo medio y los tamaños de componentes no siguen una ley de potencias, al menos, para los tamaños de red aquí estudiados. Estas características indican que la transición asociada a este ataque podría ser considerada dentro del marco de percolación explosiva [Boc+16; DSo+19]. La percolación explosiva puede corresponder a una transición discontinua, o bien, a una transición continua pero muy abrupta, caracterizada por una pendiente pronunciada en parámetro

de orden en el entorno del punto crítico. Dado que este tipo de transiciones poseen un tamaño de *crossover* típicamente grande, resulta difícil determinar el tipo de transición. Así como ha sido ampliamente discutido en revisiones recientes [DN15; DSo+19], en algunos casos la transición parece ser continua para tamaños de sistema pequeños, pero al estudiar tamaños más grandes, evidencia ser discontinua. En otros casos, ocurre lo inverso. Más aún, existen sistemas en los que la transición es discontinua, pero exhibe comportamiento crítico.

Basados en nuestros resultados, podemos afirmar que la transición asociada al ataque por RB manifiesta un comportamiento crítico, y que se aparta de la clase de universalidad correspondiente a percolación estándar. Sin embargo, nuestro análisis no resulta suficiente para realizar afirmaciones sobre la naturaleza de la transición. Para arribar a alguna conclusión al respecto, se deberían estudiar sistemas más grandes, lo cual resulta sumamente difícil dada la complejidad computacional asociada al cálculo del betweenness.

4.9. Conclusiones del capítulo

En este capítulo estudiamos transiciones de percolación inducidas por estrategias de ataque basadas en medidas de centralidad sobre redes aleatorias de tipo Erdös-Rényi. Mediante un análisis sistemático de escaleo de tamaño finito, obtuvimos tanto las posiciones de los umbrales de percolación, como los exponentes críticos que determinan la clase de universalidad de cada transición. Los umbrales de percolación calculados nos permiten verificar y, por sobre todo, cuantificar, la idea intuitiva de que las estrategias de ataque se vuelven más efectivas cuando la centralidad de los nodos es actualizada después de cada paso de remoción. En particular, pudimos observar que al mantener actualizada la información sobre la centralidad de los nodos, se puede intervenir sustancialmente sobre la transición de percolación, modificando su clase de universalidad.

Desde el punto de vista del desmantelamiento de redes, el betweenness recalculado es el ataque más efectivo, ya que es el que exhibe el umbral de percolación menor. De hecho, resulta comparable con los ataques más efectivos estudiados en la literatura [Iye+13; Bra+16; Wan+18]. Además, los exponentes asociados a este ataque son no triviales, similar a lo que ocurre en fenómenos de percolación explosiva [ADS09; FR11]. A diferencia de los ataques por grado, donde el parámetro de orden se reduce gradualmente a cero, el desmantelamiento producido por betweenness recalculado procede de manera más silenciosa, dando una imagen engañosa de integridad de la red, incluso en la proximidad de una falla catastrófica. Si consideramos infraestructuras como redes de tendido eléctrico, redes de transporte, o Internet, es razonable considerar que los nodos más cargados sean más propensos a fallar, por lo que la remoción de nodos basada en betweenness puede resultar útil no sólo para describir una estrategia de ataque, sino también interpretándola como

un mecanismo de falla. Otros autores han estudiado la vulnerabilidad de estos sistemas frente a fallas en cascada, donde la falla en nodos de alto betweenness sobrecarga otros nodos, que luego terminan fallando [Bul+10; Kor+18]. Nuestro trabajo muestra una perspectiva complementaria con la que se puede estudiar la robustez o fragilidad de este tipo de sistemas.

Capítulo 5

Ataques basados en betweenness sobre redes espaciales

5.1. Resumen

Es este capitulo analizamos la robustez ante ataques basados en la centralidad de betweenness sobre un modelo de redes espaciales aleatorias, conocido como triangulación de Delaunay. Con la idea de estudiar la importancia del caracter global o local de la centralidad utilizamos, junto con su definición original, aproximaciones del betweenness con rango acotado conocidas como ℓ-betweenness, donde todos los caminos de longitud mayor a ℓ son ignorados. Al igual que en el capitulo anterior, caracterizamos los ataques desde la perspectiva de los procesos de percolación, realizando análisis de tamaño finito y determinando los exponentes críticos. Nuestros resultados indican que para valores finitos de ℓ los ataques generan transiciones de percolación continuas que pertenecen a la clase de universalidad de percolación aleatoria. Por otra parte, los ataques por betweenness de rango completo inducen una transición discontinua, que en el límite termodinámico ocurre tras remover una fracción subextensiva de nodos. Este comportamiento es recuperado por ℓ-betweenness si la longitud ℓ escala con la longitud lineal de la red al menos como $\ell \sim L^{0,91}$. Nuestros resultados sugieren que la centralidad por betweenness codifica información en todas las escalas, de manera que no puede ser aproximada utilizando aproximaciones de rango finito sin perder eficiencia en el ataque.

5.2. Análisis cualitativo del ataque por betweenness

En la Figura 5.1 mostramos los aspectos generales del ataque por betweenness inicial. Cada uno de los paneles superiores muestra el comportamiento de uno de los observables de la transición en función de la fracción de nodos removidos. Dentro de cada panel, cada curva representa un dado tamaño lineal

del sistema. En el panel (a) de la figura presentamos el tamaño relativo de la componente gigante S_1, es decir, el parámetro de orden de la transición. Al igual que en otras transiciones, se puede observar que las curvas se vuelven más abruptas a medida que se incrementa el tamaño de la red. Dicho de otra forma, la transición se suaviza, o "redondea", para redes pequeñas, tal como vimos en el caso de las redes de Erdös-Rényi (ver Figura 4.4). No obstante, existe una sutil diferencia con respecto a las transiciones analizadas en el capítulo anterior. En aquellas, las curvas para distintos tamaños se cruzan en algún punto cercano al umbral de percolación (hay un desplazamiento hacia la derecha antes del umbral, y un desplazamiento hacia la izquierda después del mismo). En cambio, vemos en este caso que las curvas para tamaños más grandes parecen desplazarse en su totalidad hacia la izquierda. Como veremos más adelante, el análisis cuantitativo de este hecho permite deducir que la transición ocurre al inicio del ataque, es decir, $f_c = 0$.

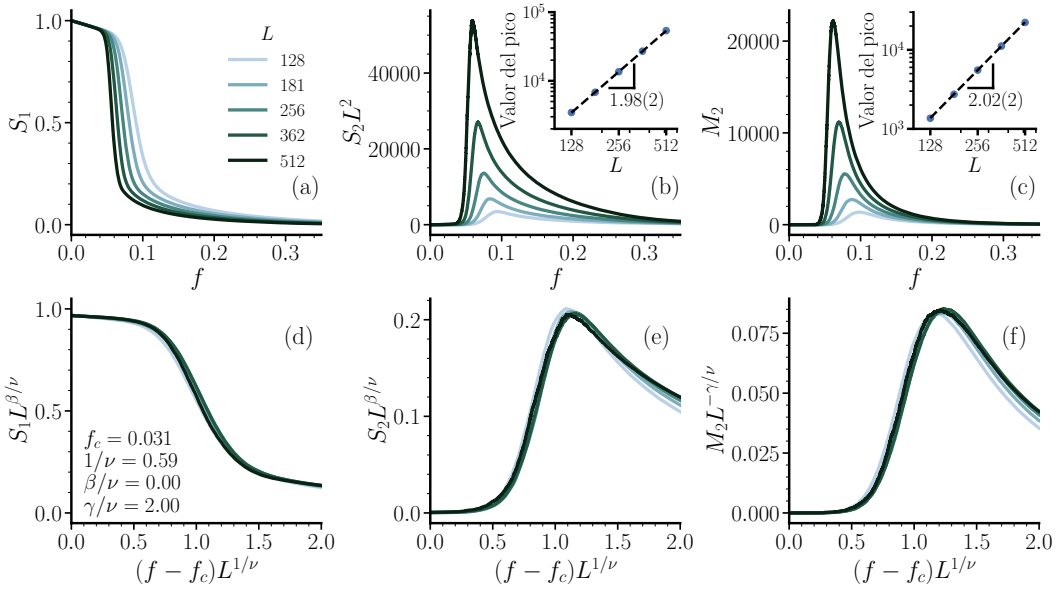

Figura 5.1: **(a-c)** Tamaño relativo de la componente gigante S_1, tamaño de la segunda componente S_2L^2 y el segundo momento de la distribución de componentes finitas M_2 como función de la fracción de nodos removidos durante el ataque por betweenness de largo alcance. Los recuadros de los paneles (b-c) muestran el escaleo de los picos de S_2L^2 y M_2 con el tamaño del sistema. Los cocientes de exponentes obtenidos con el ajuste de los picos fueron $2 - \beta/\nu = 1{,}98(2)$ y $\gamma/\nu = 2{,}02(2)$, los cuales son consistentes con los de una transición de fase de primer orden ($\beta = 0$ y $\gamma = 2\nu$). **(d-f)** Colapso de las curvas de los paneles de la izquierda basados en las ecuaciones (2.42) y (2.43).

De manera complementaria al análisis del parámetro de orden, en los paneles (b) y (c) de la figura presentamos, respectivamente, el tamaño de la segunda componente S_2L^2 y el segundo momento de la distribución de tamaños de

componentes finitas M_2, definida como

$$M_2 = \sum_s{}' s^2 n_s, \tag{5.1}$$

donde la suma primada excluye la componente gigante[1]. Para ambos observables podemos apreciar un comportamiento similar. Cada uno de ellos presenta un máximo en torno a la transición. Este pico, además, se vuelve más pronunciado y se desplaza sistemáticamente hacia la derecha en la medida que incrementamos el tamaño del sistema.

5.3. Caracterización de la transición

5.3.1. Umbral de percolación y exponentes críticos

Comenzamos la caracterización termodinámica de la transición de fase estimando algunos de sus exponentes críticos. Para ello, repetimos uno de los métodos empleados en el capítulo anterior, el cual consiste en observar el escaleo de los máximos asociados a $S_2 L^2$ y M_2. Los valores de los picos, junto con la recta de ajuste en escala log-log, se observan en los recuadros de los paneles (b-c) de la Figura 5.1. Los valores de los cocientes entre exponentes obtenidos a partir del ajuste fueron $2 - \beta/\nu = 1{,}98(2)$ y $\gamma/\nu = 2{,}02(2)$, de lo cual puede deducirse, dentro de las incertezas de la estimación, que $\beta = 0$ y $\gamma = 2\nu = d\nu$, siendo d la dimensión espacial del sistema. Observamos aquí el primer aspecto que denota que esta transición es cualitativamente distinta a las estudiadas en el Capítulo 4, ya que los exponentes obtenidos son compatibles con los de una transición de fase de primer orden [Bin81; BL84; AH10].

Como complemento del análisis anterior, estudiamos la estadística del salto máximo en la componente gigante (ver Sección 2.4.3). Los principales resultados de este análisis están resumidos en la Figura 5.2. En el panel 5.2(a) graficamos la distribución de tamaños de salto, calculada a partir de 10^4 simulaciones y para distintos tamaños. Podemos observar que las distribuciones son bimodales, con picos que se vuelven más pronunciados a medida que el tamaño del sistema aumenta. La altura del pico de la derecha no parece cambiar con L. Sin embargo, el pico derecho crece al aumentar el tamaño de la red. Las posiciones de los picos permanecen constantes en $\sim 0{,}25$ y $\sim 0{,}45$. Esto quiere decir que, en algún punto durante el transcurso del ataque, veremos un desprendimiento del 25 % o del 45 % de los nodos de la componente gigante luego de remover un sólo nodo. Este es un aspecto importante de la transición que abordaremos en la Sección 5.3.3, cuando realicemos una caracterización geométrica del ataque.

[1] Al igual que $\langle s \rangle$ (Ecuación (2.30)), M_2 puede interpretarse como medida de susceptibilidad del sistema. Para los experimentos realizados en este capítulo, verificamos que este observable permite una mejor estimación de los exponentes críticos que el tamaño medio de las componentes finitas, motivo por el cual decidimos emplearlo.

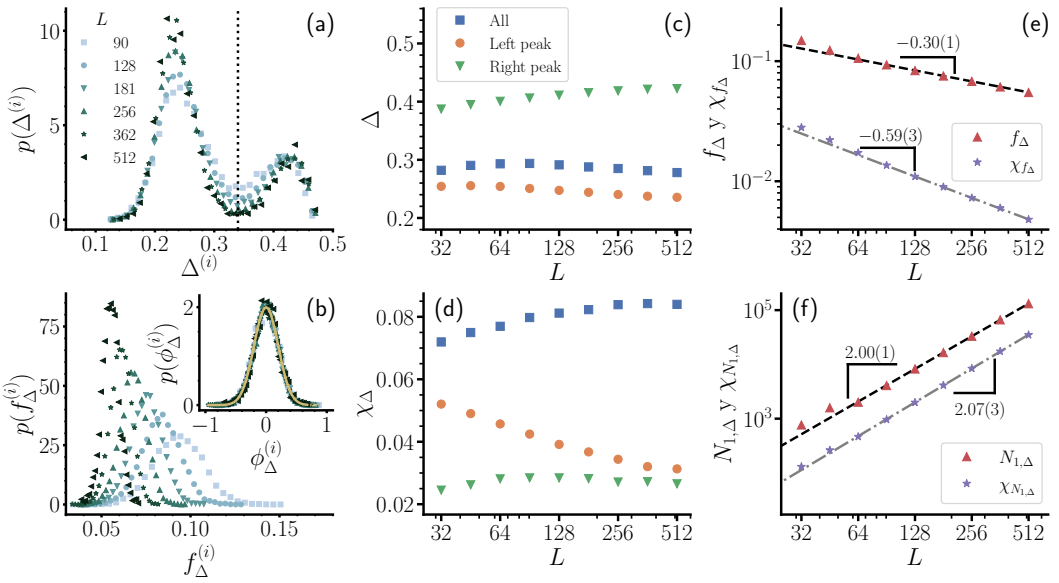

Figura 5.2: Estadística del gap, o salto máximo, para el ataque por betweenness de rango completo, en función del tamaño de la red. **(a)** Distribución de probabilidad del tamaño del gap. A medida que el tamaño de sistema aumenta la distribución se vuelve bimodal, y los valores entre los dos modos se vuelven menos probables. **(b)** Distribución de probabilidad de la posición del gap. Las curvas se mueven hacia la izquierda y se vuelven más estrechas a medida que L aumenta. Recuadro: colapso de los histogramas después de realizar el cambio de variables $\phi^{(i)} = (f_\Delta^{(i)} - f_\Delta)L^{1/\nu}$. La curva maestra (en amarillo) corresponde a una distribución gaussiana. **(c-d)** Promedio y desviación estándar del tamaño del gap. Los cuadrados azules incluyen todas las simulaciones, mientras que los círculos naranja y los triángulos verdes discriminan sólo los valores correspondientes a los picos de izquierda y derecha de la distribución de probabilidad, respectivamente. **(e)** Escaleo de la posición promedio del gap f_Δ y sus fluctuaciones χ_{f_Δ}. Analizando el escaleo de f_Δ puede observarse que el gap más grande ocurre, en el límite termodinámico, en el comienzo del ataque. A partir del escaleo de las fluctuaciones podemos estimar el exponente de la longitud de correlación $1/\nu = 0{,}59(3)$. **(f)** Promedio y fluctuaciones en el tamaño de la componente gigante en la posición del gap. El escaleo de estas dos métricas indica que la dimensión fractal de la componente gigante es $d_f = 2$, consistente con una transición primer orden.

El panel 5.2a muestra la correspondiente distribución para la posición del salto máximo. Contrario a lo que sucede con el tamaño del salto, la posición del mismo exhibe una distribución unimodal centrada. A medida que el sistema crece, las curvas se desplazan hacia la izquierda y se vuelven más pronunciadas. Basándonos en los argumentos presentados en [Fan+20], definimos la variable $\phi^{(i)} = (f_\Delta^{(i)} - f_\Delta L^{1/\nu})$, donde ν es el exponente asociado a la longitud de corre-

lación, y graficamos su distribución de probabilidad (ver recuadro del panel). Podemos observar que la distribución de esta variable se vuelve independiente del tamaño del sistema. Además, la curva maestra resultante (representada en la figura por la curva amarilla) corresponde a una gaussiana, consecuencia del teorema del límite central.

En el panel 5.2b presentamos los valores medios de saltos para distintos tamaños de sistema. Debido a la naturaleza bimodal de $p(\Delta^{(i)})$, decidimos expresar por separado los promedios computados utilizando sólo los valores asociados al pico izquierdo (círculos naranjas) y al pico derecho (triángulos verdes). La división entre cada uno de los picos corresponde a la línea gris punteada del panel 5.2a. Podemos observar que no existe una fuerte dependencia de Δ con L. Si bien el promedio del pico de la izquierda crece ligeramente y el de la derecha decrece, ambos parecen estabilizarse para los tamaños más grandes. Esto mismo puede observarse cuando se considera el promedio sobre todas las realizaciones (cuadrados azules). Como fue mencionado en la Sección 2.4.3, en una transición de percolación continua es esperable que el salto máximo promedio $\Delta(L)$ se anule en el límite termodinámico siguiendo una ley de potencia con un exponente asociado $-\beta/\nu$. Las transiciones de primer orden, en cambio, están típicamente caracterizadas por un $\beta = 0$[2], por lo que el salto máximo permanece finito incluso para $L \to \infty$. Nuestros resultados son entonces consistentes con el segundo caso, es decir, sugieren que la transición en cuestión es de primer orden.

Además de los promedios, las fluctuaciones del tamaño de salto presentan relaciones de escala similares (Ecuaciones (2.46))). Cuando tomamos todas las simulaciones, vemos un incremento de las fluctuaciones para tamaños pequeños, que luego se estabiliza hacia tamaños más grandes. Este aumento está asociado a la concentración de los valores de $\Delta^{(i)}$ hacia los picos de la distribución, la cual es más pronunciada en redes de menor tamaño.

Consideremos ahora el valor medio y las fluctuaciones asociadas a la posición del salto máximo. Como puede verse en el panel 5.2e, la relación $f_\Delta(L) - f_\Delta(\infty) \sim L^{-1/\nu_1}$ se satisface con $1/\nu_1 = 0{,}30(1)$ y $f_\Delta(\infty) = 0$. Es decir, el máximo salto en la componente gigante ocurre, en el límite termodinámico, luego de remover una cantidad subextensiva de nodos. En términos de robustez, esto indica que el ataque por betweenness es extremadamente efectivo desmantelando redes de Delaunay. En cuanto a las fluctuaciones de la posición del salto, la relación de escala $\chi_{f_\Delta}(L) \sim L^{-1/\nu}$ se satisface con $1/\nu = 0{,}59(3)$, consistente con el valor con el que fueron colapsadas las curvas en la Figura 5.1 y en el panel 5.2b.

Por último, analizamos el tamaño promedio de la componente gigante en la posición del salto máximo $N_{1,\Delta}$ y sus correspondientes fluctuaciones $\chi_{N_{1,\Delta}}$. Podemos observar en el panel 5.2f que ambas cantidades escalan con exponentes que son consistentes con $d_f = 2$. Esta relación de escala indica que la componente

[2]Una excepción a esta regla son las llamadas transiciones de percolación híbridas, las cuales describen algunos modelos de percolación explosiva [DN15].

gigante no es un fractal, como sucede en transiciones continuas, sino un objeto bidimensional. De hecho, si la relación de hiperescala $d - d_f = \beta/\nu$ se satisface, nuestro resultado indica que $\beta = 0$, validando nuevamente la hipótesis de que la transición es discontinua.

5.3.2. Distribución de tamaños de las componentes

Otro de los aspectos que estudiamos fue la distribución de tamaños de componentes cerca del umbral de percolación para el ataque por betweenness inicial. En lugar de considerar sólo las componentes de tamaño finito (como hicimos, por ejemplo, en la Sección 4.7) incluimos aquí la componente gigante. La razón de este cambio es que, de esta manera, podemos obtener mayor información sobre la naturaleza de la transición. En la Figura 5.3 mostramos histogramas construidos con 10^4 simulaciones independientes para sistemas de diferentes tamaños. El valor del parámetro de control elegido es $f = f_\Delta(L)$, por lo que depende del tamaño analizado.

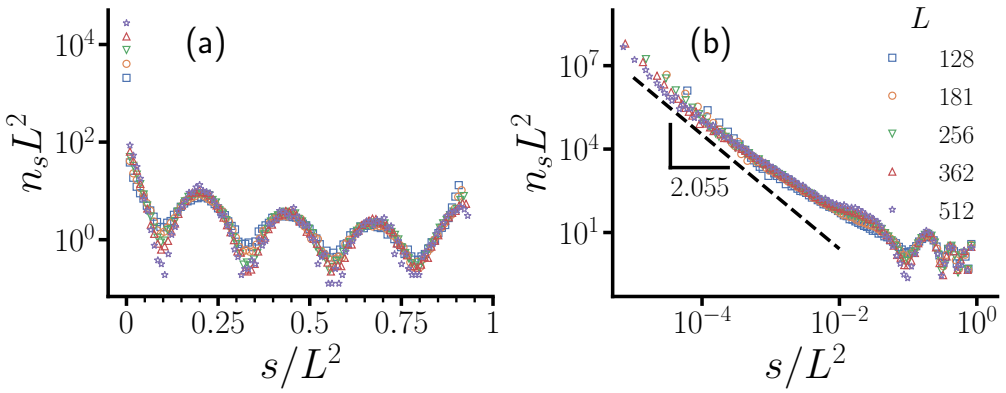

Figura 5.3: Distribución de tamaños de componentes para el ataque por betweenness inicial de largo alcance a $f = f_\Delta(L)$, incluyendo la componente gigante. Cada histograma fue construido agregando los resultados del ataque realizado sobre 10^4 redes. Las curvas están escaladas con la cantidad de nodos de la red, a fin de observar la universalidad de las mismas. **(a)** Bineado lineal, con el que puede observarse una sucesión de picos y valles que sigue una frecuencia característica. **(b)** Bineado logarítmico, el cual expone una distribución heterogénea de tamaños pequeños. La línea punteada corresponde a la distribución esperada para percolación aleatoria sobre una red regular bidimensional.

En el panel 5.3a, donde empleamos un bineado lineal, podemos observar una distribución multimodal con picos que coinciden para todos los tamaños cuando escalamos las curvas según la cantidad de nodos L^2. Podemos notar que la separación entre picos es regular, con un valor aproximado de $\sim 0{,}25$. Como veremos en la siguiente sección, tanto la existencia de los picos como su

localización puede ser entendida estudiando en detalle la evolución del ataque en realizaciones individuales. Al utilizar un bineado logarítmico, observamos que la distribución se ensancha hacia tamaños de componentes pequeños, exhibiendo un un comportamiento que asemeja una ley de potencias. A modo de referencia, incluimos en el gráfico una recta con pendiente $-2{,}055$ (línea negra discontinua), correspondiente a la distribución esperada para el proceso de percolación aleatoria sobre una red regular bidimensional.

5.3.3. Caracterización geométrica del ataque

Para ganar una mayor intuición acerca de la naturaleza de la transición, resulta útil observar una realización individual del ataque. A modo de ejemplo, en la Figura 5.4 presentamos el estado de una red i en el instante $f = f_{\Delta}^{(i)}$ para diferentes ataques. En otras palabras, presentamos la red precisamente después de que ocurra el mayor salto en la componente gigante. El tamaño de la red es de $L = 512$ y cada panel representa una estrategia de ataque diferente. Además del ataque por betweenness inicial que venimos estudiando durante el capítulo, incluimos otros dos ataques con versiones de rango acotado de betweenness (también iniciales), los cuales desarrollaremos más adelante. Para facilitar la interpretación de los resultados, mostramos sólo los nodos de la red, omitiendo los enlaces. Los puntos coloreados de los paneles superiores representan los nodos ocupados, es decir, aquellos que aún no han sido removidos de la red, donde cada color corresponde a una componente conexa distinta. En todos los casos el orden de los colores respeta el de las componentes, de acuerdo a su tamaño. En particular, la componente gigante está coloreada en azul y la segunda componente en naranja. De manera complementaria, los paneles inferiores muestran los nodos que ya han sido removidos de la red, es decir, los nodos más centrales. Si consideramos el subgrafo inducido por estos nodos y tomamos su componente gigante, obtenemos los nodos coloreados en negro. El resto de los nodos removidos corresponde a los puntos grises.

Comencemos discutiendo el panel 5.4(a), correspondiente al ataque de largo alcance. Lo primero que podemos notar es que los nodos de mayor betweenness se agrupan de dos maneras distintas sobre la red. Algunos de ellos se ubican en la zona central de la red, mientras que otros forman una serie de estrías que van del centro hacia la periferia. Debido a la simetría de la red, estas estrías son principalmente horizontales o verticales –si la red fuese circular en lugar de cuadrada, esperaríamos ver estrías radiales–. Como consecuencia de las condiciones de contorno abiertas, los nodos localizados sobre el borde se conectan entre sí mediante enlaces más largos que los del interior. Esto hace posible recorrer el exterior del grafo en pocos pasos, por lo que los nodos del borde adquieren un alto betweenness. Es gracias a estos nodos que las diferentes estrías se unen entre sí y conforman la componente gigante.

Como se puede inferir de la figura, la mayor ruptura en la red ocurre cuando se juntan dos caminos, o estrías, provenientes de bordes distintos. Como

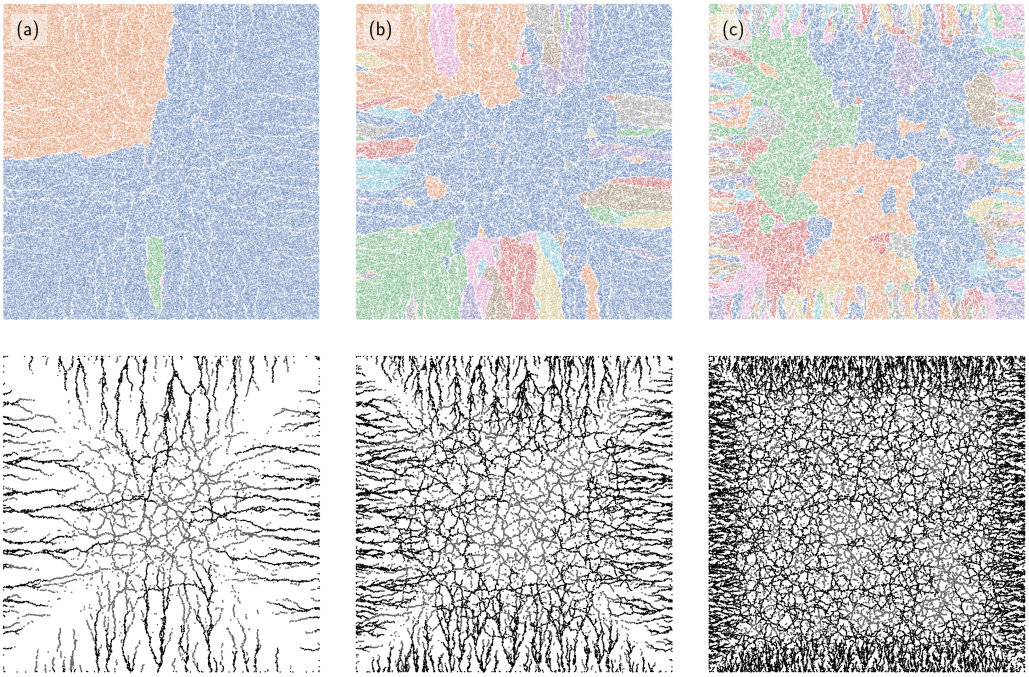

Figura 5.4: Realización individual para diferentes ataques basados en betweenness inicial sobre una red DT de tamaño $L = 512$ en el instante $f = f_\Delta^{(i)}$. Los nodos coloreados de los paneles superiores corresponden a las diferentes componentes conexas. La componente gigante y la segunda componente están coloreadas de azul y naranja, respectivamente. Los paneles inferiores muestran los nodos que ya han sido removidos de la red. Entre ellos, los representados con estrellas negras forman la componente gigante del subgrafo inducido por los nodos removidos, mientras que los puntos grises representan el resto. **(a)** Ataque con betweenness de largo alcance. Luego de remover los nodos, la red se divide en dos componentes extensivas más, eventualmente, alguna otra componente de tamaño pequeño. **(b-c)** Ataques con betweenness de rango acotado, con $\ell = 128$ y $\ell = 64$. A medida que la distancia de corte disminuye, el comportamiento de la transición cambia y emerge una distribución de tamaños de componentes más amplia, acercándose progresivamente a las distribuciones observadas para transiciones de fase continuas.

consecuencia de esto, se desprende de la red un fragmento de alrededor de un cuarto del tamaño de la misma (si los caminos provenían de bordes adyacentes) o de la mitad del tamaño (si los caminos provenían de bordes opuestos). Este hecho explica que la distribución de saltos máximos $\Delta^{(i)}$ sea bimodoal, tal como mostramos en la Figura 5.2. También nos permite interpretar los múltiples picos observados en la distribución de tamaños de componentes (5.3), los cuales están equiespaciados y ubicados aproximadamente en $s/L^2 = 1/4,\ 1/2,$ y $3/4$.

Concluimos nuestro análisis con dos observaciones acerca del ataque por

betweenness de largo alcance. Como hemos discutido hasta ahora, al remover los nodos de mayor betweenness, se produce una ruptura masiva en la red, dando lugar al desprendimiento de fragmentos de tamaño extensivo. Por otra parte, la cantidad de nodos que necesitan ser removidos para producir tal daño es en sí de carácter subextensiva. Concretamente, esta cantidad es $t_\Delta(L) = L^2 f_\Delta(L)$ que, de acuerdo con las relaciones de escala (2.46) y los resultados presentados en la Figura 5.2, escala de la forma $t_\Delta(L) \sim L^{2-1/\nu_1}$, con $\nu_1 = 0{,}30(1)$. Estas dos características nos permiten identificar a los t_Δ nodos de mayor betweenness como el *esqueleto de vulnerabilidad* o, como suele llamarse en la literatura, *vulnerability backbone* de la red. Nuestra definición es similar aunque ligeramente distinta de, por ejemplo, la definición dada en [ET10], donde se considera los nodos de mayor betweenness tales que generan una componente percolante en la red. Más allá de los detalles, nuestros resultados refuerzan los de [ET10], indicando la importancia de los nodos de alto betweenness en mantener la cohesión de las redes espaciales.

5.4. Betweenness con rango acotado

Como hemos discutido hasta ahora, la centralidad de betweenness resulta ser una métrica útil a la hora de identificar la vulnerabilidad de las redes espaciales aleatorias. Aunque hemos demostrado su efectividad, aún no nos hemos aventurado a discutir el por qué de la misma. La hipótesis más plausible que podríamos plantear es que el desempeño del ataque por betweenness está directamente asociado al carácter global de esta medida de centralidad. A diferencia de las medidas de centralidad locales, como el grado o la influencia colectiva (ver Sección 2.1.4), el betweenness de alguna manera es capaz de codificar información sobre toda la red en cada nodo. En particular, como observamos en la Sección 5.3.3, cuando se trata de una red geométrica, esta métrica nos brinda información espacial de los nodos, algo que desde el punto de vista de un ataque es de gran relevancia. Una manera en la que se puede poner a prueba esta hipótesis es generalizar la definición de betweenness incorporando un parámetro que permita ajustar la longitud de las interacciones. En esta sección desarrollaremos este método y utilizaremos, para ello, la métrica ℓ-betweenness introducida en la Sección 2.1.4.

Como primer análisis realizamos ataques sobre un ensamble de redes de Delaunay de tamaño fijo utilizando la métrica de centralidad ℓ-betweenness, tomando distintos rangos de interacción ℓ. Utilizaremos la abreviatura B para referirnos al ataque por betweenness inicial de largo alcance y Bℓ para las versiones de rango acotado. Si observamos el primer panel de la Figura 5.5, podemos verificar lo siguiente. Al utilizar una longitud de interacción mínima ($\ell = 2$) el ataque es prácticamente equivalente al ataque por grado. Como es de esperar, aumentar el parámetro ℓ hace que los ataques resultantes sean cada vez más efectivos. Eventualmente, observamos que para longitudes de interacción suficientemente grandes el ataque de rango acotado se vuelve indistinguible del

de largo alcance. Esto sugiere la existencia de una longitud de entrecruzamiento, o *crossover* $\ell^*(L)$, que podemos definir cualitativamente como la menor longitud de interacción que da lugar a un ataque igual de efectivo que el ataque por betweenness no acotado.

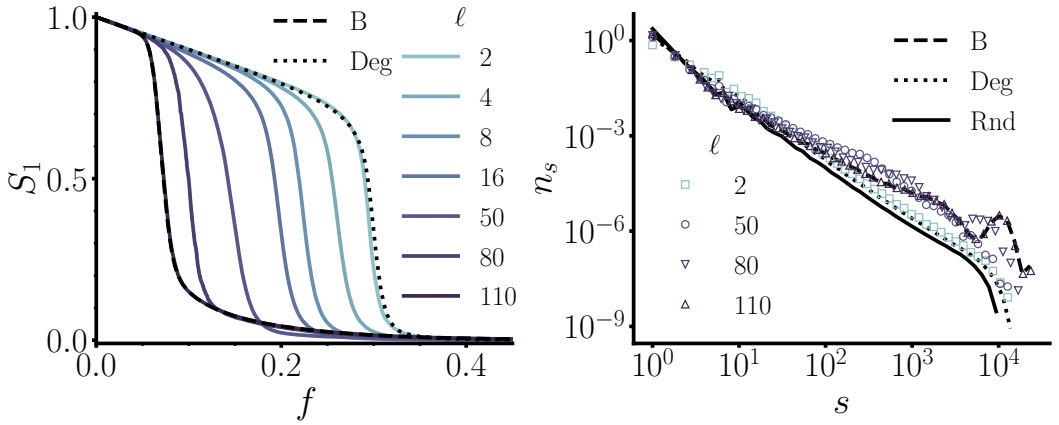

Figura 5.5: Caracterización del ataque por betweenness inicial de rango acotado, utilizando diferentes valores de corte ℓ, en redes DT con una longitud lineal $L = 256$. **(a)** Evolución del parámetro de orden S_1. Para $\ell = 2$, el ataque se comporta de manera similar al ataque basado en el grado de los nodos (linea punteada). A medida que ℓ aumenta el umbral de percolación se corre hacia la izquierda hasta que, para $\ell \approx 110$, el ataque se vuelve indistinguible de un ataque de rango completo (linea discontinua). **(b)** Distribución de tamaños de las componentes n_s en el umbral de percolación. La distribución excluye la componente gigante y el umbral de percolación está estimado como el valor de f para el cual el segundo momento de la distribución se maximiza. Para longitudes de corte pequeñas, el comportamiento es similar al de percolación aleatoria (linea sólida negra). A medida que ℓ aumenta, las curvas se vuelven más horizontales y aparece una protuberancia para valores grandes de s.

De manera alternativa, podemos comparar los ataques observando la distribución de tamaños de componentes finitas n_s en el umbral de percolación (panel 5.5(b)). Antes de discutir los resultados, es importante señalar que las cantidades graficadas en esta figura difieren de las presentadas en la Figura 5.3 en dos aspectos. En primer lugar, las distribuciones presentadas en 5.5(b) no incluyen a la componente gigante y, en segundo lugar, el estimador elegido para determinar el umbral de percolación a tamaño finito es en este caso la posición del pico del segundo momento M_2, a diferencia de la posición del máximo salto $f_\Delta(L)$ utilizada anteriormente. Para los ataques con longitud de interacción acotada, es esperable que la distribución de tamaños de componentes finitas no difiera significativamente de la de percolación aleatoria. La figura muestra que este es de hecho el caso. En cambio, se observan notorias diferencias para

valores más grandes de ℓ. Puntualmente, vemos que la distribución se aplana y que emerge un pequeño pico hacia los tamaños más grandes. Cabe destacar que, al haber excluido la componente gigante, no vemos los tres picos característicos para el ataque de largo alcance. Sin embargo, la presencia de una acumulación hacia la derecha de la distribución es un indicador de una transición abrupta, el cual actúa como un "barril de pólvora" [FL09], de manera similar a lo que sucede en algunos modelos de percolación explosiva.

La descripción anterior, de carácter más bien cualitativa, nos permite ganar intuición sobre el papel que juega la longitud de las interacciones en el desempeño de los ataques. No obstante, es posible adquirir una mayor comprensión del fenómeno estudiándolo de manera más formal. Para ello, realizamos un análisis de escala de tamaño finito con el fin de caracterizar las propiedades termodinámicas de las transiciones de percolación asociadas a cada ataque.

En primer lugar, estimamos el umbral de percolación $f_c^{\mathrm{B}\ell}$ para diferentes valores de longitud de corte, dentro del rango $2 \leq \ell \leq 16$. Utilizamos, para ello, el método de cruce definido en el capítulo anterior (ver Sección 4.4). En la Figura 5.6, análoga a la 4.3, mostramos las intersecciones de las curvas $Q(f, L) = S_1(f, L)/S_2(f, L)$ para algunos valores de ℓ. Como complemento, mostramos también las curvas correspondientes a percolación aleatoria y al ataque por grado inicial. En particular, obtenemos para percolación aleatoria el valor $f_c^{\mathrm{Rnd}} = 0{,}5002(2)$, el cual es consistente con el valor teórico $f_c = 1/2$ [SE64]. A medida que la longitud de interacción aumenta, la estimación se vuelve más difícil por dos motivos. En primer lugar, el cálculo de centralidad Bℓ se vuelve más demandante computacionalmente y, en segundo lugar, los efectos de tamaño finito se vuelven más notorios, por lo que es necesario emplear redes te mayor tamaño. Estas dos dificultades limitan nuestros cálculos a longitudes no mayores a $\ell = 16$.

De la figura podemos observar, como era de esperar, que el umbral de percolación disminuye en la medida en que la longitud de interacción aumenta. Queda en evidencia también el incremento con ℓ de la incerteza asociada a la estimación. Un aspecto interesante aquí sería determinar el valor límite al cual converge la sucesión de umbrales de percolación en la medida en que aumenta ℓ. Es decir, determinar $f_c^{\mathrm{B}\infty} := \lim_{\ell\to\infty} f_c^{\mathrm{B}\ell}$, en donde tomamos el límite $\ell \to \infty$ después de considerar el límite termodinámico $L \to \infty$. Una posibilidad sería que $f_c^{\mathrm{B}\infty} = f_c^{\mathrm{B}}$, cuyo valor determinamos anteriormente como cero. Sin embargo, no existe alguna razón fundamental por la cual estos dos valores deban coincidir. De hecho, veremos más adelante en esta sección que existen argumentos en favor de que, en realidad, la sucesión $f_c^{\mathrm{B}\ell}$ podría converger a un valor estrictamente mayor que cero.

Conjuntamente con los umbrales de percolación, estimamos algunos de los exponentes críticos de las transiciones asociadas a los ataques Bℓ. En la Figura 5.7, donde mostramos los valores de los picos de M_2 y de S_2L^2, podemos observar que no existe una variación significativa en los valores de los cocientes

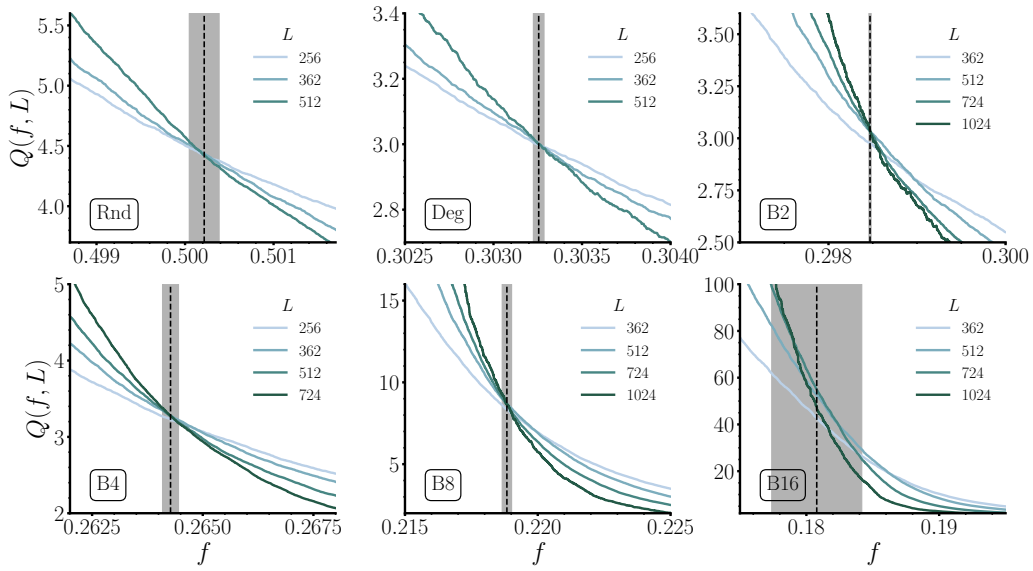

Figura 5.6: Estimación mediante el método de cruce del umbral de percolación para ataques Bℓ con diferentes longitudes de interacción ℓ. A modo de comparación, se incluye en el análisis el ataque por grado inicial y percolación aleatoria. Los promedios fueron tomados sobre sobre 10^5 simulaciones, y la relación $\ell < \ell^*(L)$ se satisface para todos los tamaños de red empleados en la estimación.

β/ν y γ/ν. Más aún, en todos los casos los valores obtenidos coinciden, dentro de las incertezas, con los valores asociados a percolación aleatoria en una red regular en dos dimensiones ($\beta/\nu = 0{,}104$ y $\gamma/\nu = 1{,}792$ [SA18]). Podemos concluir entonces que para longitudes de interacción suficientemente cortas los ataques con betweenness acotado dan lugar a transiciones continuas que pertenecen a la misma clase de universalidad que percolación aleatoria en una red regular.

Como discutimos al inicio de esta sección, el comportamiento del ataque por betweenness de alcance completo puede obtenerse utilizando la versión acotada del betweenness con una distancia de interacción lo suficientemente grande (es decir, tomando $\ell \geq \ell^*(L)$). Resulta de interés conocer el valor preciso de $\ell^*(L)$ y su variación con el tamaño de la red. Para estimar esta magnitud, calculamos la posición del salto máximo f_Δ para diferentes longitudes de interacción en sistemas de distintos tamaños (Figura 5.8). Las curvas decrecen de manera monótona cuando incrementamos ℓ hasta que alcanzan el valor correspondiente al ataque de alcance completo (líneas punteadas horizontales). Definimos entonces la longitud de *crossover* ℓ^* como el menor valor de ℓ para el cual la diferencia relativa entre $f_\Delta^{\mathrm{B}\ell}$ y f_Δ^{B} no excede un dado umbral c. El valor exacto de este umbral no es relevante, en tanto sea relativamente pequeño. Los resultados aquí mostrados corresponden a $c = 0{,}01$, y no varían significativamente si c se incrementa o disminuye en un orden de magnitud. El

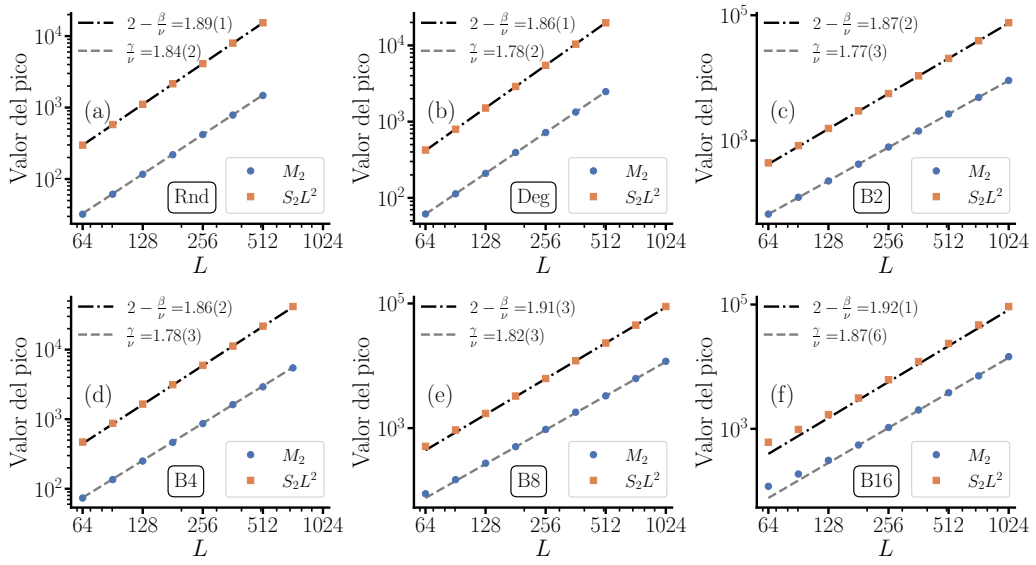

Figura 5.7: Escalado de los picos del segundo momento M_2 y la segunda componente mas grande S_2L^2 para ataques por ℓ-betweenness con diferentes valores de ℓ. Incluimos además, a modo de referencia, los resultados para percolación aleatoria y ataque inicial por grado. En todos los casos, los exponentes estimados coinciden con los asociados a percolación aleatoria en una red regular.

recuadro de la figura muestra que la distancia de *crossover* escala como una ley de potencias $\ell^* \sim L^\alpha$, con $\alpha = 0{,}91(2)$.

De manera complementaria al análisis de escala de ℓ^*, podemos extraer otra información a partir del comportamiento de $f_\Delta^{\mathrm{B}\ell}(L)$. Como puede observarse en el panel (a) de la Figura 5.8, la mayor caída en la posición del gap ocurre para longitudes de corte bajas (aproximadamente $\ell \lesssim 10$), independientemente del tamaño del sistema Sin embargo, para longitudes de interacción mayores el comportamiento de las curvas depende fuertemente del tamaño de la red. Mientras que las redes más pequeñas continúan con una caída significativa hasta que alcanzan f_Δ^{B}, los sistemas mas grandes muestran un *plateau* seguido de un punto de inflexión. Es decir, para sistemas suficientemente grandes los valores moderados de ℓ no agregan demasiada información respecto de la centralidad de los nodos. Esto podría explicarse por el hecho de que las redes de Delaunay no poseen estructuras mesoscópicas (por ejemplo, comunidades) que puedan ser censadas por una medida de centralidad de rango moderado con la finalidad de encontrar puntos débiles. Expresado de otra manera, la información requerida para desmantelar este tipo de redes aparece codificada en todas las escalas, desde los vecinos cercanos a un nodo hasta el sistema en su totalidad.

El panel 5.8b muestra las curvas escaladas por L^α y desplazadas verticalmente mediante la sustracción de $f_\Delta^{\mathrm{B}\ell}(L)$. El colapso alrededor de $\ell/L^\alpha \sim 0{,}7$ indica que el criterio adoptado para la definición de ℓ^* es apropiado, en el sentido de que el valor de ℓ^* no depende significativamente del umbral elegido.

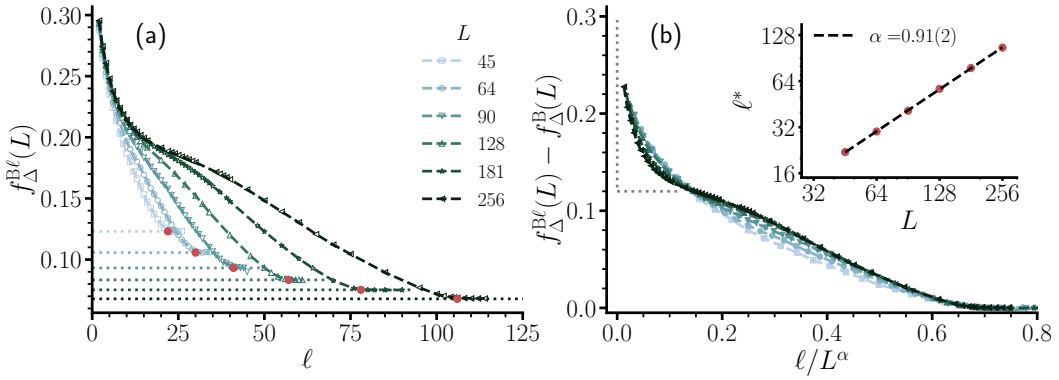

Figura 5.8: **(a)** Posición del gap $f_\Delta^{B\ell}(L)$ en función de la longitud de corte ℓ para diferentes tamaños del sistema L. Para cada valor de L, la posición del gap cae hasta que alcanza el valor correspondiente al del ataque de rango completo (lineas punteadas), el cual ocurre en $\ell = \ell^*(L)$, indicando por los puntos rojos. El escaleo de ℓ^* como función de L es graficado en el recuadro del siguiente panel. **(b)** Mismos datos que en (a) pero presentados en coordenadas escaladas. El eje vertical es desplazado mediante la sustracción de la posición del gap correspondiente al ataque de rango completo en cada tamaño, mientras que el eje horizontal es contraído por L^α.

En el límite termodinámico, este gráfico corresponde al diagrama de fases del modelo, el cual depende de los parámetros intensivos f y ℓ/L^α. En el sector inferior izquierdo del diagrama encontramos la fase percolada, mientras que en el sector opuesto aparece la fase no percolada o desmantelada. Aunque no es posible conocer la forma exacta de la curva que separa las dos fases, podemos hacernos una idea extrapolando las curvas de tamaño finito. Un primer aspecto a notar es que, por encima de un cierto tamaño, todas las curvas se cruzan en un único punto $\ell/L^\alpha = a \simeq 0{,}12$. La existencia de este punto de cruce es un indicador de que la curva de separación entre fases tiene una forma no trivial (es decir, no converge a ninguno de los ejes en el límite termodinámico). Mas aún, si nos enfocamos en la región $\ell/L^\alpha > a$, las curvas tienden a solaparse a medida que el tamaño del sistema aumenta. A partir de este hecho, podemos inferir que en el límite termodinámico las curvas no van a diferir significativamente de la curva correspondiente al sistema más grande que estudiamos, cuyo tamaño es $L = 256$. Por el contrario, para $\ell/L^\alpha < a$ las curvas parecen desplazarse sistemáticamente hacia la izquierda. Es razonable suponer que, en el límite termodinámico, la curva alcanzará el eje vertical en un valor finito f_c^*, como esquematiza la línea gris punteada de la figura. Para que esto suceda, el umbral de percolación $f_c^{B\ell}$ debería converger a un valor no nulo y en ese caso $f_c^* = f_c^{B\infty}$ (ver la discusión previa sobre la posibilidad de que $f_c^{B\infty}$ sea estrictamente mayor que cero).

Finalizamos esta sección retomando la discusión sobre la caracterización

geométrica de los ataques. En conjunto con el ataque por betweenness inicial de rango completo, en la Figura 5.4 mostramos el estado de la misma red en $f = f_{\Delta}^{\mathrm{B}\ell}(L)$, empleando ataques con longitud de interacción $\ell = 128$ (panel b) y $\ell = 64$ (panel c). Los dos valores de ℓ considerados se encuentran por debajo del valor de corte del entrecruzamiento $\ell^*(L)$, como podemos verificar extrapolando la recta del recuadro de la Figura 5.8b. Las diferencias entre estos dos ataques acotados y el ataque por betweenness completo son evidentes. Si nos fijamos en las componentes conectadas, vemos que las imágenes se corresponden con las de una percolación estándar, con una gran componente acompañada por componentes más pequeñas de tamaños heterogéneos. También la distribución espacial de los nodos removidos cambia considerablemente. Vemos que la estrías no se juntan en el centro sino que terminan a una distancia $\sim \ell$ desde el borde. Podemos utilizar esta observación para hacer una predicción heurística de ℓ^* con el siguiente razonamiento. Para que ocurra una fragmentación masiva de la red, las estrías que nacen en los distintos bordes se tienen que encontrar en el centro. Si la longitud de esas fracturas es aproximadamente ℓ, entonces $\ell^*(L) \approx L/2$. Si recordamos los valores obtenidos para los cruces del umbral de corte (ver recuadro en la Figura 5.8b) podemos ver que en el rango de tamaños estudiados, los valores están de acuerdo con esta estimación.

5.5. Discusión

En las trabajos [ET10; Erc+12], Ercsey-Ravasz, et al. reportan un estudio respecto de la contribución a la centralidad por betweenness de los caminos geodésicos de diferentes longitudes. Los autores muestran que la métrica ℓ-betweenness sigue un escalado característico respecto de ℓ que permite obtener una aproximación del betweenness de rango completo, evitando así el gran costo computacional asociado a esta medida global. Como ejemplo práctico, los autores aplican su método a la identificación de los nodos con máximo betweenness en redes con diferentes arquitecturas. En este sentido, nuestro análisis sistemático de los ataques por betweenness de rango acotado expone las limitaciones de este procedimiento. Aún cuando la mayoría de los nodos más centrales –es decir, los de mayor betweenness– son identificados correctamente mediante esta aproximación, la estadística (o termodinámica) de los ataques mediante ℓ-betweenness difiere cualitativamente de la del ataque completo. Esto sugiere que la interacción topológica de los nodos en la escala global provee una contribución no despreciable al betweenness, que se pierde cuando se realizan aproximaciones locales. También notamos que el algoritmo introducido por estos autores no reduce la complejidad computacional requerida en el cálculo de las diferentes centralidades por ℓ-betweenness (en el Apéndice D analizamos en detalle los tiempos de cómputo requeridos).

En la literatura de la física estadística las transiciones de fase en la triangulación de Delaunay (DT) se han estudiado extensamente [JV95; Lim+00; DAF16; McC87]. En conjunto con la dilución de conexiones y el reconecta-

do local, DT representa un modelo de nodos embebidos espacialmente con desorden topológico o congelado [JW04; BV14]. Una de las consecuencias de este desorden es el efecto de suavizado o redondeo en las transiciones de fase. En algunos sistemas, el desorden destruye la coexistencia de fases y tiene un efecto de redondeo en las transiciones de primer orden [Car99]. Llamativamente, para el caso particular del ataque de rango completo por betweenness ocurre lo opuesto. Si consideramos una red regular con condiciones de contorno periódicas, entonces todos los nodos son equivalentes, por lo que cualquier ataque basado en centralidades iniciales es equivalente a una percolación por sitios aleatoria. Sin embargo, el desorden asociado a las redes DT tornan las transiciones mas abruptas, transformando la transición de segundo orden a una de primer orden[3].

A la hora de pensar en una posible aplicación de nuestro trabajo a redes reales, conviene recordar que las redes DT han sido utilizadas en el modelado de redes viales en ciudades [Kir+18], donde la centralidad por betweenness tiene una interpretación directa en términos de flujo de tráfico. En este contexto, la eliminación de un nodo de alto betweenness puede interpretarse como la saturación de una intersección de mucho tráfico. Es importante observar que tanto la discontinuidad de la transición de percolación como su localización en $f = 0$ son indicadores de la potencial fragilidad de este tipo de sistemas. De hecho, se han reportado recientemente transiciones de percolación discontinuas en redes de tráfico reales [Zen+20]. Todos estos elementos nos lleva a pensar que nuestro trabajo puede ser útil para entender y predecir congestiones de transito en grandes ciudades.

Para finalizar la discusión queremos enfatizar que todos los ataques considerados en este capitulo son iniciales. Es decir, las medidas de centralidad son calculadas al comienzo del ataque, sin ser actualizadas posteriormente cuando se remueven los nodos. Una manera de interpretar este tipo de ataques es considerar que los mismos ocurren en una escala temporal en la cual la remoción de un nodo ocurre de manera suficientemente rápida, de modo que la medida de centralidad no requiere de un ajuste. La extensión de nuestros análisis a ataques cuyas medidas se actualizan puede ser una linea interesante de estudio para llevar a cabo en el futuro. Sería útil además comparar dichos resultados con los actuales. En este sentido, conviene recordar que los resultados del capítulo anterior muestran que en las redes de Erdös-Rényi el hecho de recalcular el betweenness cambia la clase de universalidad de la transición y reduce el umbral de percolación. Como el umbral de percolación para el ataque por betweenness inicial en las redes de Delaunay ya es óptimo, este aspecto no podría mejorar con el recalculado de la centralidad. De hecho, el umbral de percolación fue determinado en $f_c = 0$ para la versión recalculada del betweenness [NMH16]. Sin embargo, otros aspectos termodinámicos de la

[3]Si bien en esta tesis analizamos sólo redes con condiciones de contorno abiertas, resultados preliminares muestran que el comportamiento cualitativo de las transiciones no cambia al emplear condiciones periódicas.

transición, como por ejemplo sus exponentes críticos, sí podrían cambiar con el recalculado de la centralidad.

5.6. Conclusiones del capítulo

En este capítulo estudiamos el proceso de desmantelamiento de la triangulación de Delaunay en dos dimensiones, mediante la remoción de nodos basada en la centralidad por betweenness. Estudiamos la fragmentación de esta red desde la perspectiva de un proceso de percolación, caracterizando el orden y la criticalidad de la transición por medio de un análisis de escala de tamaño finito. Además del betweenness en su definición estándar, empleamos aproximaciones al mismo que se conocen como ℓ-betweenness. En esta aproximación se ignoran los caminos mas largos que ℓ, de manera que se puede cambiar de una aproximación local a una global incrementando el valor de ℓ.

Encontramos que los ataques basados en betweenness de rango completo generan una transición discontinua que ocurre a $f_c^{\mathrm{B}} = 0$. Por otra parte, valores finitos de ℓ producen transiciones continuas con valores $f_c^{\mathrm{B}\ell} > 0$ que pertenecen a la clase de universalidad de percolación aleatoria en redes regulares de dos dimensiones. Mediante la variación sistemática del parámetro ℓ determinamos que el comportamiento del rango completo se recupera cuando ℓ se incrementa con el tamaño lineal del sistema L siguiendo la relación $\ell \sim L^{\alpha}$, donde α toma un valor cercano a 1. Esto sugiere que cualquier aproximación finita del betweenness empeora la efectividad del ataque, cambiando la transición de forma cuantitativa (incrementando el umbral de percolación) y cualitativa (modificando el orden de la transición).

Capítulo 6

Conclusiones

En la primera parte de esta tesis realizamos una caracterización estadística de la base de datos de partidas de ajedrez más extensa disponible, en su momento, en el mundo. A partir de esta base de datos construimos dos tipos de redes de jugadores de ajedrez, la primera integrada por partidas jugadas sobre tableros físicos (OTB) y la segunda con partidas desarrolladas en portales de internet (Portal). En estas redes los nodos representan a los jugadores y las partidas jugadas establecen la relación entre los nodos. Una vez construidas, efectuamos una caracterización detallada de la topología de las mismas, con el enfoque puesto en evaluar la relación entre la estructura de la red y el desempeño de los jugadores, utilizando el nivel de Elo de los jugadores como atributo. Para la caracterización estructural utilizamos métricas existentes e introdujimos una nueva medida que evalúa el fenómeno de club de ricos tomando como medida el Elo de los jugadores. La relación entre atributos y estructura fue estudiada principalmente evaluando correlaciones entre la estructura de comunidades y el Elo de los jugadores.

En la segunda parte de este trabajo investigamos la robustez ante fallas y ataques dirigidos de dos tipos de redes sintéticas obtenidas a partir de modelos generativos. Una de estas clases corresponde al modelo de grafos aleatorios de Erdös-Rényi (ER) y la otra, a las redes planares de Delaunay (DT) construidas a partir de un criterio de proximidad. Para caracterizar la robustez de ambas redes utilizamos ataques dirigidos sobre nodos guiados por medidas de centralidad. Empleamos medidas de centralidad locales y globales, y consideramos ataques en sus versiones iniciales y recalculadas. Para evaluar la efectividad de los ataques y realizar un estudio comparativo de los mismos utilizamos las herramientas de análisis de los fenómenos críticos. Específicamente, estudiamos las transiciones de percolación inducidas por los ataques, realizando extensas simulaciones numéricas y análisis de tamaño finito para encontrar los exponentes críticos que determinan las clases de universalidad de las transiciones asociadas. A continuación resumimos los principales resultados y contribuciones de esta tesis en cada una de las principales áreas de estudio.

Comenzamos con las redes de jugadores de ajedrez. Los primeros resultados

se refieren a un análisis estadístico comparativo entre las bases de datos de jugadores OTB y Portal. La base OTB recopila principalmente partidas de jugadores profesionales, con escasos jugadores aficionados. Esto es de algún modo esperable ya que la actividad de los jugadores aficionados sobre tablero se desarrolla en contextos informales donde los juegos no son registrados. A raíz de este sesgo, la base OTB tiene jugadores con Elo mayores que Portal. Además, el Elo de los jugadores de OTB evoluciona más lentamente, pues los jugadores tienen posiciones más consolidadas. En cambio, en los portales de internet todas la partidas son almacenadas, independientemente del Elo de los jugadores. Los jugadores en este caso exhiben una variabilidad mayor en el Elo que el de los jugadores sobre tablero y también un puntaje medio comparativamente menor.

En cuanto a las redes construidas a partir de los datos, observamos que ambas presentan características en común, pero que existen marcadas diferencias entre ellas. Las dos redes tienen redes una distribución de grado asimétrica de cola larga, siendo la asimetría más notoria en la red Portal, la cual se asemeja a una ley de potencias con un exponente cercano a 1,5. Una de las principales diferencias que observamos entre las redes está relacionada a las correlaciones entre los grados de los nodos. OTB es una red asortativa, al igual que muchas redes sociales [New02]. En cambio, Portal tiene una estructura disortativa similar a la observada en algunas redes tecnológicas y biológicas. Otro resultado que se relaciona con esta diferencia es que la correlación entre el Elo y la actividad de los jugadores es marcadamente positiva en OTB mientras que en Portal es débil e incluso negativa. Es decir, en Portal los jugadores de mayor Elo no son necesariamente los mas activos, algo que sí sucede en OTB.

Observamos que en los dos modos de juego (Internet y tablero físico) los jugadores de élite tienden a formar grupos densamente conectados. Cuantificamos este aspecto introduciendo un nuevo coeficiente de club de ricos que selecciona a los nodos de acuerdo a su Elo (y que puede ser extendido a cualquier otro atributo ordinal de los nodos). En la variante estándar de este coeficiente, donde se seleccionan a los nodos por su grado, sólo la red OTB exhibe este comportamiento. Esta diferencia es por un lado un indicio del efecto que tienen los algoritmos que forman las parejas de jugadores en los portales de internet, y por otra parte refleja el hecho de que los jugadores más activos en Portal son los de nivel medio.

En el análisis de la transitividad observamos que ambas redes presentan un valor decreciente del coeficiente de clustering como función del grado de los nodos, aunque con un decrecimiento más leve en OTB que en Portal. En particular, en esta última el comportamiento se aproxima a la relación $c(k) \sim k^{-1}$, indicador de que la red de Portal tiene una estructura jerárquica más pronunciada [Rav+02; RB03; Rav04] .

Por ultimo, los resultados sobre la detección de comunidades y las correlaciones con el nivel de Elo representan uno de los aportes más originales de esta tesis. Para este análisis utilizamos varios algoritmos para la detección de comunidades

en las redes OTB y Portal, en particular uno muy reciente—al momento del análisis—propuesto por Newman y Clauset [NC16], que puede detectar correlaciones entre el Elo de los jugadores y las estructuras de comunidades. Todos estos análisis se contrastaron con experimentos realizados en modelos nulos apropiados, que aleatorizan las redes preservando las distribuciones de grado. Utilizando los algoritmos estándar encontramos que el Elo de cada comunidad muestra una dispersión menor que la que se observa cuando se considera a la red en su totalidad. Este es un primer indicio de la existencia de correlaciones entre el Elo de los jugadores y las comunidades. Además, el algoritmo de Newman y Clauset confirma que existe una correlación significativa entre la estructura de comunidades y el desempeño de los jugadores, es decir, su Elo.

Una de las principales conclusiones que se extraen de los resultados aquí obtenidos es que, en las redes de ajedrez, el Elo es el factor que determina la dinámica con la que las redes se conforman. Las dos redes analizadas tienen soportes diferentes que imponen determinadas restricciones y características. Por un lado, tenemos partidas llevadas a cabo sobre tableros físicos, que sólo pueden llevarse a cabo si los dos oponentes se encuentran en el mismo lugar. Por otro lado, Internet abre la puerta a la interacción a distancia, o globalizada, de los jugadores. Por este motivo, OTB mantiene muchas de las características que frecuentemente se observan en redes sociales, mientras que Portal presenta una mezcla entre algunas de estas características y otras vinculadas a las redes tecnológicas. Estas diferencias desaparecen cuando las métricas calculadas tienen en cuenta el Elo de los jugadores (por ejemplo, en el caso de asortatividad, correlaciones, club de ricos y comunidades). Esto indica que una misma dinámica de red, en este caso representada por la distribución de partidas entre jugadores, puede tener lugar en redes con topologías diferentes. Dicho de otro modo, la inclusión de atributos en la descripción de un sistema complejo como una red compleja es en algunos casos fundamental para lograr un entendimiento completo sobre el mismo.

En los estudios concernientes a la robustez de redes frente a ataques dirigidos utilizamos en primer lugar las redes aleatorias del modelo de Erdös-Rényi como banco de pruebas. Estudiamos las transiciones de percolación inducidas por estrategias de ataques que se basan en medidas de centralidad locales y globales; tanto en sus versiones iniciales como recalculadas. Mediante un análisis sistemático de escaleo de tamaño finito, obtuvimos las posiciones de los umbrales de percolación y los exponentes críticos que determinan la clase de universalidad de cada transición. Los umbrales de percolación calculados nos permitieron cuantificar la efectividad de los ataques y confirmar la idea intuitiva de que estrategias de ataques son más efectivas cuando la centralidad de los nodos es actualizada después de cada paso de remoción. Un resultado a destacar es que al mantener actualizada la información sobre la centralidad de los nodos, se puede cambiar la naturaleza de la transición de percolación, modificando su clase de universalidad. Nuestros resultados muestran que el betweenness recalculado es el ataque más efectivo para desmantelar las redes de ER, ya

que tiene el umbral de percolación menor de todos los ataques que ensayamos. Además, es comparable con las estrategias de desmantelamiento más efectivas que se han reportado en la literatura [Iye+13; Bra+16; Wan+18]. Por otra parte, los exponentes asociados a este ataque son no triviales, en semejanza a lo que ocurre en fenómenos de percolación explosiva [ADS09; FR11]. A diferencia de los ataques por grado, donde el parámetro de orden se reduce gradualmente a cero, el desmantelamiento producido por betweenness recalculado procede de manera suave, dando una imagen engañosa de integridad de la red, incluso en la proximidad de una falla catastrófica.

Los resultados que obtuvimos tienen consecuencias prácticas ya que el betweenness de un nodo se puede interpretar como una medida de la carga que el mismo soporta. En infraestructuras tales como redes de tendido eléctrico, redes de transporte, o Internet, es razonable considerar que los nodos más cargados sean los más propensos a fallar, por lo que los ataques guiados por betweenness pueden ser interpretados como un mecanismo de falla en este contexto. Este aspecto ha sido explorado por otros autores que han estudiado la vulnerabilidad de este tipo de sistemas frente a fallas en cascada. Cuando se produce una falla en un nodo de alto betweenness, el flujo se reconfigura y sobrecarga a otros nodos del sistema, que también terminan fallando [Bul+10; Kor+18]. Es decir se produce naturalmente un recalculado de la carga de los nodos. En este sentido, los resultados obtenidos en las redes de ER aportan información sobre la fragilidad de ciertos sistemas de infraestructura.

Continuando con los estudios de robustez ante ataques dirigidos analizamos los procesos de desmantelamiento en triangulaciones de Delaunay en dos dimensiones. Nos enfocamos en ataques iniciales donde la remoción de nodos está basada en la centralidad por betweenness. Además de utilizar el betweenness en su definición estándar empleamos la aproximación ℓ-betweenness, en la cual se omiten los caminos con distancia mayor a ℓ, de manera que los ataques están guiados por una medida de centralidad local cuando ℓ es pequeño y se transforman en globales a medida que este parámetro aumenta. Al igual que en las redes de ER, analizamos los resultados desde la perspectiva de un proceso de percolación, caracterizando el orden y la criticalidad de la transición por medio de un análisis de escala de tamaño finito. Encontramos que los ataques basados en betweenness de rango completo generan una transición discontinua que ocurre cuando la fracción crítica de nodos removidos es cero en el limite termodinámico ($f_c^{\mathrm{B}} = 0$). Por otra parte, los ataques por betweenness de rango acotado que utilizan valores finitos de ℓ producen transiciones continuas con valores $f_c^{\mathrm{B}\ell} > 0$ que pertenecen a la clase de universalidad de la percolación por sitios aleatoria en redes regulares en dos dimensiones.

Otro resultado que consideramos importante lo obtuvimos analizando sistemáticamente el efecto de la variación del parámetro ℓ. En particular, encontramos que el comportamiento del ataque de rango completo se recupera cuando ℓ se incrementa con el tamaño lineal del sistema L siguiendo la relación $\ell \sim L^{\alpha}$, donde α toma un valor cercano a 1. Este resultado sugiere que cualquier

aproximación finita en el calculo del betweenness empeora la efectividad del ataque guiado por esta medida. Los resultados en la red DT muestran que la transición de percolación cambia cuantitativa y cualitativamente. Esto es, el umbral de percolación aumenta y se modifica el orden de la transición.

Por último, es interesante destacar que la detección de comunidades y la evaluación de la robustez de redes no son problemas desconectados. Trabajos relativamente recientes muestran que ataques dirigidos que tienen en cuenta la estructura de comunidades de redes complejas son particularmente eficientes para afectar la funcionalidad del sistema asociado a la red [PB12; RGG15; Yua+16]. El criterio en este caso consiste en detectar la estructura de comunidades de la red y para fragmentarla eliminar las conexiones entre comunidades [RGG15]. Como es de esperar, este mecanismo de ataque ha demostrado ser particularmente efectivo en redes con comunidades bien definidas, aunque también ha mostrado dar buenos resultados en redes con una débil estructura de comunidades. El análisis estadístico de las transiciones de percolación asociadas a este tipo de ataques no ha sido estudiado aún.

Nuestro trabajo se puede extender en varias direcciones. En primer lugar, es importante mencionar que el Elo es un sistema que se utiliza en otros deportes, algunos de ellos digitales. Sería interesante estudiar las redes asociadas a estos deportes y evaluar hasta qué punto el Elo determina la conformación de las mismas. En particular, un análisis similar al desarrollado aquí fue hecho para el popular deporte de tablero Go en [LZ21].

En cuanto al tema de robustez en redes, un tema de interés consiste en determinar las condiciones que un ataque debe tener para dar lugar a transiciones de fase no triviales (transiciones discontinuas o con exponentes críticos diferentes a los de percolación aleatoria). En el presente trabajo, observamos que tales transiciones son posibles con ataques iniciales y con ataques recalculados, dependiendo del sustrato o red que se esté estudiando. En los casos aquí estudiadas observamos también que es necesario que la medida de centralidad sea global para poder modificar cualitativamente la transición. Naturalmente, nos podríamos preguntar si esta condición es siempre necesaria. En otras palabras, el interrogante que aparece es si es posible obtener una transición de percolación explosiva empleando un ataque de alcance local.

Para finalizar, otra dirección posible a seguir es la de determinar de manera analítica (o semi-analítica) los umbrales de percolación y los exponentes críticos asociados para cada uno de los ataques aquí estudiados. Este aspecto es de interés más bien teórico y, fuera de los casos simples como los ataques por grado, se vuelve extremadamente difícil de abordar.

Apéndice A

Sistema de puntuación Elo

Luego de varios intentos, en el año 1970 la Federación Internacional de Ajedrez (FIDE, por sus siglas en francés) adoptó el sistema de puntuación que se denomina Elo para calificar el nivel de juego de sus jugadores. Este sistema estadístico tomó el nombre del físico y ajedrecista aficionado Árpád Élő, quien fue su creador. En el desarrollo de su sistema Élő observó que la distribución de rendimientos de los jugadores de ajedrez se asemejaba a una distribución normal, por lo cual la diferencia entre las distribuciones de rendimiento de dos jugadores es también una distribución normal, sólo que más dispersa [Gli95]. De manera que el sistema de calificación de los jugadores desarrollado por Árpád Élő se encuadra dentro del área de la estadística en modelado de *comparación de pares*.

En el sistema de Elo cada jugador posee un puntaje numérico el cual es estimado a partir de victorias, derrotas y empates en enfrentamientos contra otros jugadores. Además, la diferencia entre los Elos de dos jugadores permite estimar el resultado esperado del partido. Si un jugador A de Elo R_A se enfrenta a un jugador B con Elo R_B, las expectativas (E_i con $i = A$ o B) que tienen los jugadores de obtener una dada puntuación se calculan de la siguiente manera,

$$E_A = \frac{1}{1 + 10^{(R_B - R_A)/400}}$$

$$E_B = \frac{1}{1 + 10^{(R_A - R_B)/400}},$$

se debe tener en cuenta que en una partida un jugador obtiene 1 punto si gana, $\frac{1}{2}$ si empata y 0 si pierde. A modo de ejemplo, una diferencia de 200 puntos implica que el jugador de mayor Elo (supongamos el jugador A) tiene una expectativa de 0,75 puntos, este resultado expresado en términos de la probabilidad de victoria P_v, de empate P_e y de derrota P_d queda:

$$E_A = 1 \times P_v + \frac{1}{2} \times P_e + 0 \times P_d = P_v + \frac{1}{2} P_e.$$

Este sistema de calificación tiene además la ventaja de que se actualiza mediante un algoritmo simple, que utiliza los resultados que los jugadores

Categoría	Abreviatura	Rango de Elo
Gran Maestro Internacional	IG	> 2500
Maestro Internacional	IM	$2400 - 2500$
Maestro FIDE	FM	$2300 - 2400$
Maestro Candidato	CM	$2200 - 2300$
Experto	E	$2000 - 2200$
Clase A	CA	$1800 - 2000$
Clase B	CB	$1600 - 1800$
Clase C	CC	$1400 - 1600$
Clase D	CD	$1200 - 1400$

Cuadro A.1: Categorías de jugadores de ajedrez según rango de Elo

obtienen en un torneo. Si un jugador supera el puntaje esperado, su Elo aumenta. En caso contrario, el Elo disminuye. Las actualizaciones de los Elos de los jugadores se realizan de manera incremental y utilizan un parámetro, llamado factor K, el cual depende de la categoría, siendo $K = 16$ Elo para jugadores de alto nivel y $K = 32$ Elo para jugadores menos expertos. Si en una partida un jugador A posee un puntaje esperado de E_A puntos, y luego del juego obtuvo S_A, su actualización de Elo será

$$R'_A = R_A + K \cdot (S_A - E_A). \tag{A.1}$$

Si bien la escala de puntuaciones tiene como límite mínimo al cero, el Elo máximo no está limitado, aunque es muy poco probable que un jugador exceda los 3000 Elo. Actualmente, los jugadores de ajedrez poseen Elos menores a 2900 [1], sin embargo los mejores software de ajedrez como *Stockfish* [TK22] ya superan los 3500 Elo.

El sistema de puntuación de Elo es utilizado por FIDE y USCF para clasificar tanto los torneos como los jugadores en categorías. Los jugadores suelen ser clasificados según la Tabla A.1. Por atra parte, La FIDE clasifica los torneos considerando el promedio de Elo de los jugadores. Las categorías cambian cada 25 puntos, comenzando con la categoría 1 con Elos de 2251 a 2275, hasta la categoría 22 con Elos superiores a 2776 para los hombres, y las mismas categorías para el caso de las mujeres pero con 200 puntos menos, por lo tanto la correspondiente categoría 1 sería de 2051 hasta 2075. Por otra parte la Federación de Ajedrez Estadounidense clasifica a los jugadores en 14 categorías según su Elo, desde la categoría A (Elos de 100 a 199) hasta la categoría Senior Master (Elo 2400 ó superior) en incrementos de 200.

[1] Magnus Carlsen, actual campeón mundial, se propone superar esta marca.

Apéndice B

Procesamiento de las bases de datos de ajedrez

El formato más frecuentemente empleado para almacenar partidas de ajedrez es el formato PGN. Un archivo PGN puede contener una o varias partidas, e incluye tanto los metadatos (jugadores, lugar y fecha de la partida, resultado, etc.), como la secuencia de movimientos, expresada en notación algebraica (único sistema de notación oficial admitido por la FIDE). A modo de ejemplo, la célebre partida disputada entre Kasparov y Topalov jugada en Wijk aan Zee en 1999 se expresa en formato PGN como

```
[Event "It (cat.17)"]
[Site "Wijk aan Zee (Netherlands)"]
[Date "1999.??.??"]
[Round "?"]
[White "Garry Kasparov"]
[Black "Veselin Topalov"]
[Result "1-0"]

1. e4 d6 2. d4 Nf6 3. Nc3 g6 4. Be3 Bg7 5. Qd2 c6 6. f3 b5 7. Nge2 Nbd7
8. Bh6 Bxh6 9. Qxh6 Bb7 10. a3 e5 11. O-O-O Qe7 12. Kb1 a6 13. Nc1 O-O-O
14. Nb3 exd4 15. Rxd4 c5 16. Rd1 Nb6 17. g3 Kb8 18. Na5 Ba8 19. Bh3 d5
20. Qf4+ Ka7 21. Rhe1 d4 22. Nd5 Nbxd5 23. exd5 Qd6 24. Rxd4 cxd4 25. Re7+
Kb6 26. Qxd4+ Kxa5 27. b4+ Ka4 28. Qc3 Qxd5 29. Ra7 Bb7 30. Rxb7 Qc4
31. Qxf6 Kxa3 32. Qxa6+ Kxb4 33. c3+ Kxc3 34. Qa1+ Kd2 35. Qb2+ Kd1 36. Bf1
Rd2 37. Rd7 Rxd7 38. Bxc4 bxc4 39. Qxh8 Rd3 40. Qa8 c3 41. Qa4+ Ke1 42. f4
f5 43. Kc1 Rd2 44. Qa7 1-0
```

Nótese que algunos campos contienen información faltante, expresada con signos de pregunta.

Las dos bases de datos empleadas en esta tesis (OTB y Portal) fueron recibidas en archivos PGN, uno por cada base. Los archivos fueron procesados en `python` empleando la librería `chess` [Fie22], con la cual extrajimos los metadatos relevantes (no fue necesario, para este trabajo, utilizar la secuencia de movimientos).

White	Black	WhiteElo	BlackElo	Result	PlyCount	Date	Site
Badongski	SLuciano	1373	1441	1-0	33	2015.09.01	FICS freechess.org
Muttakin	tuffshaq	1805	1692	1-0	70	2015.09.01	FICS freechess.org
dabearsfan	thatgirl	1814	1924	0-1	132	2015.09.01	FICS freechess.org
BobCashFlow	Lodi	1665	1640	0-1	40	2015.09.01	FICS freechess.org
DragonWasp	surios	1852	1647	1-0	31	2015.09.01	FICS freechess.org

Cuadro B.1: Ejemplo del esquema tabular de las bases de datos de ajedrez.

Una vez extraídos los metadatos, pasamos a un formato tabular del estilo de la Tabla B utilizando la librería `pandas` y los almacenamos en archivos CSV. A partir de estos archivos, construimos y analizamos los grafos empleando diferentes librerías y código propio.

Apéndice C

Métodos de función generatriz aplicados al problema de percolación

C.1. Percolación en el modelo de configuración

Consideremos el modelo de configuración (Sección 2.2.2) para grafos con distribución de grado p_k. Tomemos un nodo cualquiera de la red, y sigamos uno de sus enlaces (asumiendo que tiene al menos uno) hasta el nodo ubicado en el otro extremo. Definamos a u como la probabilidad de que el nodo del extremo no pertenezca a la componente gigante. En el modelo de configuración esta probabilidad no depende del nodo elegido, ni del enlace que decidimos seguir.

Para pertenecer a la componente gigante, un nodo tiene que estar conectado a dicha componente a través de uno de sus vecinos. De manera equivalente, un nodo no pertenece a la componente gigante si y sólo si no está conectado a dicha componente a través de ninguno de sus vecinos, lo cual sucede con probabilidad u^k si tiene k vecinos[1]. Promediando sobre toda la red, la probabilidad de que un nodo cualquiera no pertenezca a la componente gigante es $\sum_k p_k u^k$, donde p_k es la distribución de grado. Dada su gran utilidad, esta suma merece su propia notación,

$$g_0(u) = \sum_k p_k u^k. \tag{C.1}$$

La función $g_0(u)$ se conoce como *función generadora de probabilidad* o *función generatriz* para la distribución p_k [Wil05]. Como $g_0(u)$ es la probabilidad media de que un nodo no pertenezca a la componente gigante, entonces,

$$S = 1 - g_0(u) \tag{C.2}$$

[1]En este paso estamos asumiendo que las probabilidades de conexión son todas independientes, lo cual no sería cierto si, por ejemplo, existiesen conexiones entre los vecinos. En el límite de redes grandes y esparsas, sin embargo, el modelo de configuración da lugar a redes que son localmente árboles, por lo que este tipo de conexiones entre vecinos no existen.

es la probabilidad de que sí pertenezca a la componente gigante o, de forma equivalente, S es la fracción de nodos de la red que pertenece a dicha componente. Vemos entonces que para obtener S es necesario calcular el valor de u. Para realizar este cálculo consideremos lo siguiente. La probabilidad de que un nodo v no esté conectado a la componente gigante a través de uno de sus vecinos en particular (llamemos a este vecino w) es igual a la probabilidad de que w no pertenezca a la componente gigante a través de ninguno de sus otros vecinos. Si w tiene otros k vecinos (es decir, w tiene $k+1$ vecinos si contamos a v), entonces esta probabilidad es nuevamente u^k. Basta ahora promediar sobre toda la red, pero debemos ser cautelosos en este punto. Como estamos hablando de vecinos de un nodo, ahora la variable k no está distribuida según la distribución de grado p_k, sino que sigue la distribución de grado en exceso q_k (Ecuación (2.3)). Pero, por definición, esta probabilidad es u, con lo cual llegamos a la ecuación autoconsistente

$$u = \sum_k q_k u^k. \tag{C.3}$$

En forma análoga a la definición (C.1), la suma anterior es la función generatriz asociada a q_k y se denota como

$$g_1(u) = \sum_k q_k u^k. \tag{C.4}$$

Expresando en términos de funciones generatrices, llegamos entonces a que

$$u = g_1(u). \tag{C.5}$$

En conjunto, (C.2) y (C.5) dan la solución para el tamaño S de la componente gigante. Aunque en general no es posible encontrar una solución cerrada para estas ecuaciones, es posible resolverlas de forma numérica partiendo de la distribución de grado. Pero incluso si no conocemos la distribución de grado exactamente, podemos tener una idea general sobre el comportamiento de u como veremos a continuación. Primero, notemos que $g_1(1) = \sum_k q_k = 1$, ya que q_k es una distribución de probabilidad debidamente normalizada. Luego, $u = g_1(u)$ tiene siempre una solución trivial para $u = 1$, sin importar la distribución de grado, que corresponde a la situación en la cual no existe componente gigante. Sin embargo, la ecuación (C.5) puede tener otra solución que sí de lugar a una componente gigante.

Como $g_1(u)$ es una serie de potencias con coeficientes no negativos, sus derivadas son también no negativas cuando $u \geq 0$, por lo que $g_1(u)$ es una función creciente con concavidad hacia arriba. Dado que toma el valor 1 cuando $u = 1$, su gráfica debe ser cualitativamente similar a una de las curvas de la Figura C.1. Las soluciones a la ecuación $u = g_1(u)$ vienen dadas por los puntos de intersección entre $y = g_1(u)$ a recta $y = u$ (línea punteada en la figura).

La solución trivial $u = 1$ aparece en la esquina superior derecha de la figura y está siempre presente, pero puede haber o no una segunda solución para

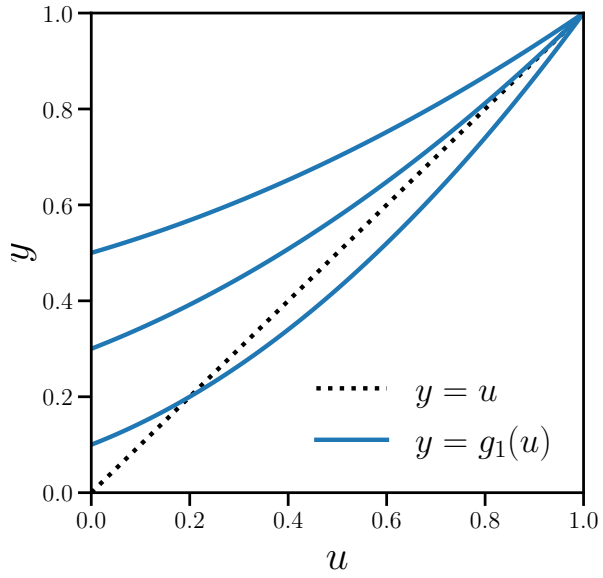

Figura C.1: Solución gráfica de la Ecuación (C.5). Las soluciones a la ecuación $u = g_1(u)$ están dadas por los puntos para los cuales $y = g_1(u)$ intercepta a la recta $y = u$.

$u < 1$, dependiendo de la forma que tome $g_1(u)$. En particular, podemos ver que esta solución no trivial aparece justo cuando la pendiente de $g_1(u)$ en $u = 1$ es igual a 1. Es decir, cuando se cumple

$$g_1'(1) > 1. \tag{C.6}$$

Utilizando las definiciones (C.4) y (2.3), obtenemos

$$g_1'(1) = \sum_k k q_k = \frac{1}{\langle k \rangle} \sum_k k(k+1) p_k = \frac{\langle k \rangle^2 - \langle k \rangle}{\langle k \rangle} > 1, \tag{C.7}$$

por lo que la solución no trivial aparece cuando

$$\begin{aligned} \langle k \rangle^2 - 2\langle k \rangle &> 0 \\ \kappa &> 2, \end{aligned} \tag{C.8}$$

donde $\kappa = \langle k^2 \rangle / \langle k \rangle$ es el parámetro de heterogeneidad de la red.

C.2. Ataque por grado inicial

Una vez introducida la noción del uso de funciones generatrices para abordar el problema de percolación aleatoria en redes complejas, pasaremos a extender el análisis a ataques dirigidos. En particular, en esta sección detallaremos el enfoque introducido por Callaway, et al. en [Cal+00].

Consideremos un proceso de percolación arbitrario sobre una red con distribución de probabilidad p_k en el cual se ha removido una fracción f de nodos. Definamos r_k como la probabilidad condicional de que un nodo esté ocupado (es decir, que no haya sido removido) dado que tiene grado k. De acuerdo con esta definición, $p_k r_k$ es la probabilidad conjunta de que un nodo esté ocupado y tenga grado k. Asociada a esta probabilidad podemos definir la función

$$F_0(x) = \sum_{k=0}^{\infty} p_k r_k x^k. \tag{C.9}$$

Esta función se asemeja a una función generatriz, pero no lo es exactamente ya que no está adecuadamente normalizada. Concretamente, verificamos fácilmente que $F_0(1) = \sum_k p_k r_k = 1 - f$.

Si elegimos un enlace al azar y lo seguimos en alguna dirección, el nodo al que llegamos tiene una distribución de grado en exceso proporcional a kp_k (Ecuación (2.3)), en lugar de p_k. Luego, la función equivalente a (C.9) para tal vértice es

$$F_1(x) = \frac{\sum_k k p_k r_k x^{k-1}}{\sum_k k p_k} = \frac{F_0'(x)}{\langle k \rangle}. \tag{C.10}$$

Consideremos ahora la función generatriz $H_1(x)$ asociada a la probabilidad de que un enlace elegido uniformemente al azar conduzca a una componente con una dada cantidad de nodos ocupados. Esta componente podría tener cero nodos si el vértice al que llegamos siguiendo el enlace fue removido, lo cual ocurre con probabilidad $1 - F_1(1)$, o podría pasar que el enlace conduzca a un nodo con otros k enlaces salientes, distribuidos de acuerdo con $F_1(x)$. Esto significa que $H_1(x)$ satisface la ecuación autoconsistente [Cal+00]

$$H_1(x) = 1 - F_1(1) + x F_1[H_1(x)]. \tag{C.11}$$

De manera similar se puede obtener la función generatriz $H_0(x)$ asociada a la distribución de probabilidad para el tamaño de componente a la cual pertenece un nodo escogido al azar, la cual satisface

$$H_0(x) = 1 - F_0(1) + x F_0[H_1(x)]. \tag{C.12}$$

Juntas, las ecuaciones (C.9)-(C.11) determinan la distribución de tamaños de componentes para un proceso de percolación por sitios generalizado en una red con distribución de grado arbitraria. A partir de ellas es posible obtener muchas cantidades de interés. Por ejemplo, la fracción de nodos que pertenece a la componente gigante viene dada por

$$S_1 = 1 - H_0(1) = F_0(1) - F_0(u), \tag{C.13}$$

donde u es una solución de la ecuación autoconsistente

$$u = 1 - F_1(1) + F_1(u). \tag{C.14}$$

La expresión anterior es similar a la Ecuación (C.5). Por un lado, $u = 1$ es la solución correspondiente a la ausencia de componente gigante ($S_1 = 0$). Además, $F_1(x)$ es una serie de potencias con coeficientes no negativos al igual que $g_1(x)$, por lo que es creciente y con concavidad hacia arriba en el intervalo $0 \leq u \leq 1$, lo cual significa que la solución no trivial (C.14) aparece cuando

$$F_1'(1) = 1. \tag{C.15}$$

Especificando la distribución de grado p_k y definiendo una estrategia de ataque r_k, el valor de $F_1'(1)$ depende exclusivamente de la fracción f de nodos removidos. La fracción exacta para la cual se satisface (C.15) es, por definición, el umbral de percolación f_c.

Otra cantidad de interés que podemos obtener a partir de H_0 es el tamaño medio de las componentes finitas $\langle s \rangle$, la cual viene dada por la ecuación

$$\langle s \rangle = H_0'(1). \tag{C.16}$$

De hecho, es posible verificar que $\langle s \rangle$ diverge cuando se cumple la igualdad (C.15).

Para el caso particular del ataque por grado inicial, la probabilidad condicional r_k para un dado valor de f toma la forma

$$r_k = \begin{cases} 1, & f \leq P_{k+1} \\ \dfrac{P_k - f}{p_k}, & P_{k+1} < f < P_k \\ 0, & f \geq P_k \end{cases} \tag{C.17}$$

Podemos entender esta expresión de la siguiente manera. Si la fracción de nodos removidos es suficientemente chica, todos los nodos de grado k están aún en la red. En cambio, si f es suficientemente grande, significa que todos los nodos de grado k han sido removidos. Para un rango intermedio de f, sólo una fracción (que varía linealmente con f) de los nodos de grado k fue removida.

A partir de (C.17) podemos resolver numéricamente las ecuaciones (C.13) y (C.14) para calcular el parámetro de orden y la susceptibilidad a cada valor del parámetro de control f. Además, podemos estimar el umbral de percolación estimando el valor de f para el cual se alcanza la igualdad (C.15).

C.3. Ataque por grado recalculado

En [KKG20], los autores Kim, et al. encontraron una forma de obtener, de manera semi-analítica, un valor preciso para el punto de percolación en el ataque por grado recalculado. Su estrategia fue calcular cómo varía la distribución de grado al remover una fracción f de nodos mediante este ataque, y emplear la teoría de funciones generatrices descripta en la Sección C.1.

Supongamos que conocemos la distribución de grado $p_k(f)$ al momento en que una fracción f de los nodos fue removida. La distribución $p(f + df)$ se

puede obtener de la siguiente manera. En primer lugar, en la medida en que f se incrementa por df, p_0 aumenta en df y p_K decrece en df, donde K es el máximo grado de la red (el cual irá disminuyendo en la medida en la que prosigamos con el ataque). Al mismo tiempo, cuando un nodo de grado K es removido, los grados de sus vecinos disminuyen en 1. La probabilidad de que un nodo de grado k esté conectado a un dado enlace de uno de los nodos desactivados es la probabilidad de grado en exceso $kp_k(f)/\langle k\rangle(f)$. Combinando estos dos efectos, obtenemos un sistema completo de relaciones de recurrencia[2]

$$\begin{aligned}
p_0(f+df) &= p_0(f) + K\frac{p_1}{\langle k\rangle}df + df \\
p_1(f+df) &= p_1(f) + K\frac{2p_2 - p_1}{\langle k\rangle}df \\
&\vdots \\
p_k(f+df) &= p_k(f) + K\frac{(k+1)p_{k+1} - kp_k}{\langle k\rangle}df \\
&\vdots \\
p_K(f+df) &= p_K(f) + K\frac{-Kp_K}{\langle k\rangle}df - df
\end{aligned} \tag{C.18}$$

El lado derecho de cada ecuación contiene el término de ganancia debido a los nodos de grado $(k+1)$ conectados con el nodo removido y los términos de pérdida debido a los nodos de grado k conectados a dicho nodo. Las únicas excepciones son $k = 0$ y $k = K$, donde sólo tenemos ganancia o pérdida, respectivamente. Notemos también que, en teoría, K no necesariamente es finito. Sin embargo, si la distribución de grado original es homogénea, es posible introducir un valor máximo finito para K sin perder precisión. En [KKG20], los autores definen K como el mayor k tal que $p_k(f) > 10^{-10}$ y en esta tesis emplearemos el mismo criterio. Por poner un ejemplo, para una red de Erdös-Rényi con $\langle k\rangle = 3{,}5$, este criterio equivale a $K = 21$.

[2]Omitimos la dependencia en f en algunos de los términos para simplificar la notación.

Apéndice D

Estudio comparativo de la complejidad computacional de los algoritmos para el cálculo de betweenness de rango acotado

Como mencionamos en la Sección 2.1.4, el algoritmo *naïve* para el cálculo del betweenness tiene una complejidad computacional de $\mathcal{O}(N^2M)$. El algoritmo de Brandes [Bra01] reduce esta complejidad a $\mathcal{O}(NM)$, la mejor cota que ha sido obtenida hasta el momento. Como este algoritmo está basado en una búsqueda en anchura, o *breadth-first search*, puede ser adaptado fácilmente para calcular el betweenness de rango acotado. Concretamente, basta con detener cada búsqueda cuando se alcanza la longitud de interacción ℓ deseada. En caso de que se requiera calcular el ℓ-betweenness para múltiples longitudes, debemos correr el algoritmo desde el inicio para cada valor de ℓ. Luego, si quisiéramos hacer un análisis comparativo para todas las longitudes $\ell \leq \ell^*$, la complejidad total sería $\mathcal{O}(\ell^* MN)$.

En [ET10], los autores proponen un algoritmo alternativo para el cómputo del betweenness de rango acotado en el cual se fija un valor ℓ^* y se calcula, de manera simultánea, la centralidad ℓ-betweenness para todas las longitudes $\ell \leq \ell^*$. Los autores muestran que la complejidad computacional $\mathcal{C}$ de este algoritmo está acotada por $\mathcal{O}(NM) < \mathcal{C} < \mathcal{O}(\ell^* MN)$.

Para comparar el tiempo de cómputo de ambos algoritmos realizamos el siguiente experimento. Elegimos implementaciones de cada algoritmo en lenguaje `C` (utilizamos una implementación del algoritmo de Ercsey-Ravasz proveída por sus autores, mientras que para el de Brandes acotado escogimos la implementación de la librería `igraph`). Para una dada red de tamaño L calculamos todos los valores de ℓ-betweenness hasta el valor $\ell^*(L)$, estimado de acuerdo con la relación de escala obtenida en la Sección 5.4. Repetimos este cáclculo diez veces y promediamos los tiempos de ejecución. Luego, repetimos este procedimiento para redes de diferentes tamaños. Como ℓ^* escala con el

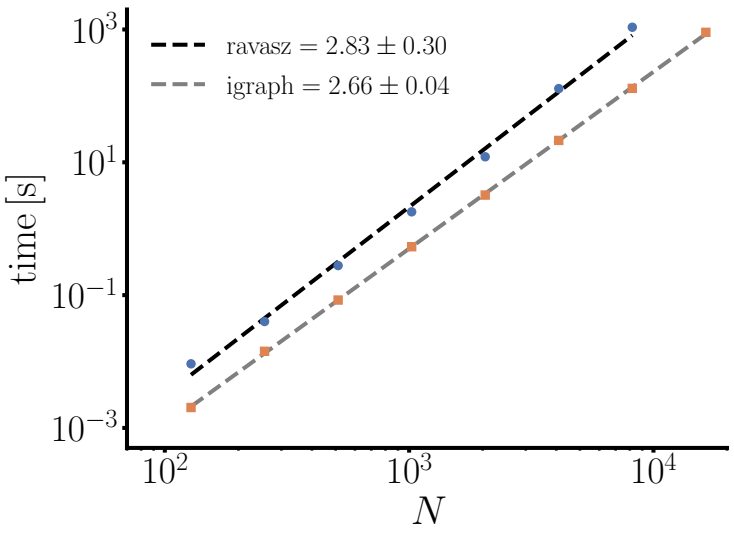

Figura D.1: Comparación entre la complejidad computacional de dos algoritmos para el cálculo de betweenness con rango acotado. Los círculos azules representan el algoritmo de Ercsey-Ravasz, mientras que los cuadrados naranjas corresponden al algoritmo de Brandes acotado implementado en la librería `igraph`.

tamaño de la red como $\ell^* \sim L^\alpha = N^{\alpha/2}$ para redes DT y como $M \sim N$, entonces la complejidad para el peor caso en ambos algoritmos es $\mathcal{O}(N^{2+\alpha/2}) \approx \mathcal{O}(N^{2,46})$, de acuerdo con la estimación de α obtenida en 5.4.

Como podemos apreciar en la Figura D.1, la estimación numérica para la complejidad de los dos algoritmos está cerca de de la predicción teórica. En particular, el algoritmo de Ercsey-Ravasz es ligeramente más lento, indicando que la complejidad computacional promedio está cerca del límite superior, como mencionan los autores en [ET10]. Nuestra prueba de desempeño indica entonces que no existe ventaja alguna en términos de tiempo de cómputo en utilizar el algoritmo de Ercsey-Ravasz para estimar ℓ^*. De hecho, emplear el algoritmo estándar de Brandes tiene tiene la ventaja adicional de que es posible omitir ciertos valores de ℓ en la estimación de ℓ^*, reduciendo aún más el tiempo de cómputo requerido.

Bibliografía

[AAN04] Réka Albert, István Albert y Gary L. Nakarado. «Structural vulnerability of the North American power grid». En: *Physical Review E* 69.2 2 (2004), págs. 1-4. ISSN: 1063651X. DOI: 10.1103/PhysRevE.69.025103.

[AB02] Réka Albert y Albert-László Barabási. «Statistical mechanics of complex networks». En: *Rev. Mod. Phys.* 74 (1 ene. de 2002), págs. 47-97. DOI: 10.1103/RevModPhys.74.47.

[ABP20] Nahuel Almeira, Orlando Vito Billoni y Juan Ignacio Perotti. «Scaling of percolation transitions on Erdös-Rényi networks under centrality-based attacks». En: *Physical Review E* 012306 (2020), págs. 1-9. DOI: 10.1103/PhysRevE.101.012306.

[ADS09] Dimitris Achlioptas, Raissa M. D'Souza y Joel Spencer. «Explosive Percolation in Random Networks». En: *Science* 323.5920 (mar. de 2009), págs. 1453-1455. DOI: 10.1126/science.1167782.

[AE16] Mathieu Acher y François Esnault. «Large-scale Analysis of Chess Games with Chess Engines: A Preliminary Report». En: *arXiv preprint arXiv:1607.04186* (2016).

[AH10] N. A. M. Araújo y H. J. Herrmann. «Explosive Percolation via Control of the Largest Cluster». En: *Physical Review Letters* 105.3 (jul. de 2010). DOI: 10.1103/physrevlett.105.035701.

[AJB00] Réka Albert, Hawoong Jeong y Albert-László Barabási. «Error and attack tolerance of complex networks». En: *Nature* 406.6794 (jul. de 2000), págs. 378-382. DOI: 10.1038/35019019.

[Alm+17] Nahuel Almeira y col. «Structure constrained by metadata in networks of chess players». En: *Scientific reports* 7.1 (2017), págs. 1-10. DOI: 10.1038/s41598-017-15428-z.

[Alm+21] Nahuel Almeira y col. «Explosive dismantling of two-dimensional random lattices under betweenness centrality attacks». En: *Chaos, Solitons & Fractals* 153 (dic. de 2021), pág. 111529. DOI: 10.1016/j.chaos.2021.111529.

[Alv+19] L. G. Alvarez-Zuzek y col. «Dynamic vaccination in partially overlapped multiplex network». En: *Phys. Rev. E* 99 (1 ene. de 2019), pág. 012302. DOI: 10.1103/PhysRevE.99.012302.

[Ant71] J.M. Anthonisse. *The Rush in a Directed Graph*. Stichting Mathematisch Centrum. Mathematische Besliskunde, 1971. URL: https://books.google.com.ar/books?id=tVH-rQEACAAJ.

[ASV16] Arshia Atashpendar, Tanja Schilling y Thomas Voigtmann. «Sequencing chess». En: *EPL (Europhysics Letters)* 116.1 (2016), pág. 10009. DOI: 10.1209/0295-5075/116/10009.

[Bar11] M. Barthélemy. «Spatial networks». En: *Physics Reports* 499.1-3 (2011), págs. 1-101. ISSN: 03701573. DOI: 10.1016/j.physrep.2010.11.002.

[Bar99] A. Barabasi. «Emergence of Scaling in Random Networks». En: *Science* 286.5439 (oct. de 1999), págs. 509-512. DOI: 10.1126/science.286.5439.509.

[Bas+14] N. Bastas y col. «Method for estimating critical exponents in percolation processes with low sampling». En: *Physical Review E* 90.6 (dic. de 2014). DOI: 10.1103/physreve.90.062101.

[BC78] Edward A Bender y E.Rodney Canfield. «The asymptotic number of labeled graphs with given degree sequences». En: *Journal of Combinatorial Theory, Series A* 24.3 (mayo de 1978), págs. 296-307. DOI: 10.1016/0097-3165(78)90059-6.

[BE06] Stephen P. Borgatti y Martin G. Everett. «A Graph-theoretic perspective on centrality». En: *Social Networks* 28.4 (oct. de 2006), págs. 466-484. DOI: 10.1016/j.socnet.2005.11.005.

[Bin81] K. Binder. «Finite size scaling analysis of ising model block distribution functions». En: *Zeitschrift für Physik B Condensed Matter* 43.2 (jun. de 1981), págs. 119-140. DOI: 10.1007/bf01293604.

[BL84] K. Binder y D. P. Landau. «Finite-size scaling at first-order phase transitions». En: *Physical Review B* 30.3 (ago. de 1984), págs. 1477-1485. DOI: 10.1103/physrevb.30.1477.

[Blo+08] Vincent D Blondel y col. «Fast unfolding of communities in large networks». En: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10 (oct. de 2008), P10008. DOI: 10.1088/1742-5468/2008/10/p10008.

[Boc+16] S. Boccaletti y col. «Explosive transitions in complex networks' structure and dynamics: Percolation and synchronization». En: *Physics Reports* 660 (nov. de 2016), págs. 1-94. DOI: 10.1016/j.physrep.2016.10.004.

[BPV04] M. Boguña, R. Pastor-Satorras y A. Vespignani. «Cut-offs and finite size effects in scale-free networks». En: *The European Physical Journal B - Condensed Matter* 38.2 (mar. de 2004), págs. 205-209. DOI: 10.1140/epjb/e2004-00038-8.

[BR06] Béla Bollobás y Oliver Riordan. «The critical probability for random Voronoi percolation in the plane is 1/2». En: *Probability Theory and Related Fields* 136.3 (2006), págs. 417-468. ISSN: 01788051. DOI: 10.1007/s00440-005-0490-z.

[Bra+08] U. Brandes y col. «On Modularity Clustering». En: *IEEE Transactions on Knowledge and Data Engineering* 20.2 (feb. de 2008), págs. 172-188. DOI: 10.1109/tkde.2007.190689.

[Bra+16] Alfredo Braunstein y col. «Network dismantling». En: *Proceedings of the National Academy of Sciences* 113.44 (oct. de 2016), págs. 12368-12373. DOI: 10.1073/pnas.1605083113.

[Bra01] Ulrik Brandes. «A faster algorithm for betweenness centrality». En: *The Journal of Mathematical Sociology* 25.2 (jun. de 2001), págs. 163-177. DOI: 10.1080/0022250x.2001.9990249.

[Bra96] Ĭ. Brankov. *Introduction to Finite-size Scaling*. Leuven notes in mathematical and theoretical physics: Mathematical physics. Leuven University Press, 1996. ISBN: 9789061867586. URL: https://books.google.com.ar/books?id=mIjMAAAACAAJ.

[BT09] Bernd Blasius y Ralf Tönjes. «Zipf's Law in the Popularity Distribution of Chess Openings». En: *Phys. Rev. Lett.* 103 (21 nov. de 2009), pág. 218701. DOI: 10.1103/PhysRevLett.103.218701.

[Bul+10] Sergey V. Buldyrev y col. «Catastrophic cascade of failures in interdependent networks». En: *Nature* 464.7291 (abr. de 2010), págs. 1025-1028. DOI: 10.1038/nature08932.

[BV14] Hatem Barghathi y Thomas Vojta. «Phase Transitions on Random Lattices: How Random is Topological Disorder?» En: *Physical Review Letters* 113.12 (sep. de 2014). DOI: 10.1103/physrevlett.113.120602.

[BZ09] A. M. Becker y R. M. Ziff. «Percolation thresholds on two-dimensional Voronoi networks and Delaunay triangulations». En: *Physical Review E* 80.4 (oct. de 2009), págs. 1-9. ISSN: 15393755. DOI: 10.1103/PhysRevE.80.041101.

[Cal+00] Duncan S. Callaway y col. «Network Robustness and Fragility: Percolation on Random Graphs». En: *Physical Review Letters* 85.25 (dic. de 2000), págs. 5468-5471. DOI: 10.1103/physrevlett.85.5468.

[Car99] John Cardy. «Quenched randomness at first-order transitions». En: *Physica A: Statistical Mechanics and its Applications* 263.1-4 (1999), págs. 215-221. ISSN: 03784371. DOI: 10.1016/S0378-4371(98)00489-0.

[CG11] Philippe Chassy y Fernand Gobet. «Measuring Chess Experts' Single-Use Sequence Knowledge: An Archival Study of Departure from 'Theoretical' Openings». En: *PLoS ONE* 6.11 (2011), e26692. DOI: 10.1371/journal.pone.0026692.

[CG17] Bruno Requião da Cunha y Sebastián Gonçalves. «Performance of attack strategies on modular networks». En: *Journal of Complex Networks* 5.6 (jun. de 2017), págs. 913-923. DOI: 10.1093/comnet/cnx015.

[Cha11] Rick Chartrand. «Numerical Differentiation of Noisy, Nonsmooth Data». En: *ISRN Applied Mathematics* 2011 (mayo de 2011), págs. 1-11. DOI: 10.5402/2011/164564.

[Cin19] Matteo Cinelli. «Generalized rich-club ordering in networks». En: *Journal of Complex Networks* 7.5 (feb. de 2019). Ed. por Luciano Costa, págs. 702-719. DOI: 10.1093/comnet/cnz002.

[CNM04] Aaron Clauset, M. E. J. Newman y Cristopher Moore. «Finding community structure in very large networks». En: *Physical Review E* 70.6 (dic. de 2004), pág. 066111. DOI: 10.1103/physreve.70.066111.

[Coh+00] Reuven Cohen y col. «Resilience of the Internet to Random Breakdowns». En: *Physical Review Letters* 85.21 (nov. de 2000), págs. 4626-4628. DOI: 10.1103/physrevlett.85.4626.

[Coh+01] Reuven Cohen y col. «Breakdown of the Internet under Intentional Attack». En: *Physical Review Letters* 86.16 (abr. de 2001), págs. 3682-3685. DOI: 10.1103/physrevlett.86.3682.

[Col+06] V. Colizza y col. «Detecting rich-club ordering in complex networks». En: *Nat Phys* 2.2 (ene. de 2006), págs. 110-115. DOI: 10.1038/nphys209.

[Cru+04] Paolo Crucitti y col. «Error and attack tolerance of complex networks». En: *Physica A: Statistical Mechanics and its Applications* 340.1-3 (2004), págs. 388-394. ISSN: 03784371. DOI: 10.1016/j.physa.2004.04.031.

[CZD12] W. Chen, Z. Zheng y R. M. D'Souza. «Deriving an underlying mechanism for discontinuous percolation». En: *EPL (Europhysics Letters)* 100.6 (dic. de 2012), pág. 66006. DOI: 10.1209/0295-5075/100/66006.

[Da +14] R. A. Da Costa y col. «Solution of the explosive percolation quest: Scaling functions and critical exponents». En: *Physical Review E* 90.2 (2014), págs. 1-13. ISSN: 15502376. DOI: 10.1103/PhysRevE.90.022145.

[DAF16] Marcelo M. De Oliveira, Sidiney G. Alves y Silvio C. Ferreira. «Continuous and discontinuous absorbing-state phase transitions on Voronoi-Delaunay random lattices». En: *Physical Review E* 93.1 (2016), págs. 1-7. ISSN: 24700053. DOI: 10.1103/PhysRevE.93.012110.

[Din+14] Binbin Ding y col. «Numerical analysis of percolation cluster size distribution in two-dimensional and three-dimensional lattices». En: *European Physical Journal B* 87.8 (2014). ISSN: 14346036. DOI: 10.1140/epjb/e2014-40996-4.

[DN15] Raissa M. D'Souza y Jan Nagler. «Anomalous critical and supercritical phenomena in explosive percolation». En: *Nature Physics* 11.7 (jul. de 2015), págs. 531-538. DOI: 10.1038/nphys3378.

[DSo+19] Raissa M. D'Souza y col. «Explosive phenomena in complex networks». En: *Advances in Physics* 68.3 (jul. de 2019), págs. 123-223. DOI: 10.1080/00018732.2019.1650450.

[Dua+12] Xujun Duan y col. «Large-scale brain networks in board game experts: Insights from a domain-related task and task-free resting state». En: *PLoS ONE* 7.3 (2012), págs. 1-11. ISSN: 19326203. DOI: 10.1371/journal.pone.0032532.

[ER59] Paul Erdős y A. Rényi. «On random graphs». En: *Publicationes Mathematicae (Debrecen)* 6 (1959), pág. 290.

[ER60] Paul Erdos y Alfred Renyi. «On the evolution of random graphs». En: *Publ. Math. Inst. Hungary. Acad. Sci.* 5 (1960), págs. 17-61.

[Erc+12] Mária Ercsey-Ravasz y col. «Range-limited centrality measures in complex networks». En: *Physical Review E* 85.6 (jun. de 2012). DOI: 10.1103/physreve.85.066103.

[ET10] Mária Ercsey-Ravasz y Zoltán Toroczkai. «Centrality Scaling in Large Networks». En: *Physical Review Letters* 105.3 (jul. de 2010). DOI: 10.1103/physrevlett.105.038701.

[Fag07] Giorgio Fagiolo. «Clustering in complex directed networks». En: *Phys. Rev. E* 76 (2 ago. de 2007), pág. 026107. DOI: 10.1103/PhysRevE.76.026107.

[Fan+20] Jingfang Fan y col. «Universal gap scaling in percolation». En: *Nature Physics* 16.4 (feb. de 2020), págs. 455-461. DOI: 10.1038/s41567-019-0783-2.

[FH16] Santo Fortunato y Darko Hric. «Community detection in networks: A user guide». En: *Physics Reports* 659 (2016), págs. 1-42. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2016.09.002.

[Fie22] Niklas Fiekas. *python-chess: a chess library for Python*. https://python-chess.readthedocs.io/en/latest/. [Accedido el 28 de Marzo de 20222]. 2022.

[FL09] Eric J. Friedman y Adam S. Landsberg. «Construction and Analysis of Random Networks with Explosive Percolation». En: *Physical Review Letters* 103.25 (dic. de 2009). DOI: 10.1103/physrevlett.103.255701.

[For10] Santo Fortunato. «Community detection in graphs». En: *Physics Reports* 486.3-5 (feb. de 2010), págs. 75-174. DOI: 10.1016/j.physrep.2009.11.002.

[Fos+11] David V Foster y col. «Clustering drives assortativity and community structure in ensembles of networks». En: *Physical Review E* 84.6 (2011), pág. 066117. DOI: 10.1103/PhysRevE.84.066117.

[FR11] S. Fortunato y F. Radicchi. «Explosive percolation in graphs». En: *Journal of Physics: Conference Series* 297 (2011), pág. 012009. DOI: 10.1088/1742-6596/297/1/012009.

[Fre04] Linton C Freeman. *The development of social network analysis*. Booksurge Publishing, jul. de 2004.

[Fre77] Linton C. Freeman. «A Set of Measures of Centrality Based on Betweenness». En: *Sociometry* 40.1 (mar. de 1977), pág. 35. DOI: 10.2307/3033543.

[Gli95] Mark E. Glickman. «Chess rating systems». En: *American Chess Journal* 3 (1995), págs. 59-102.

[Gra+11] Peter Grassberger y col. «Explosive Percolation is Continuous, but with Unusual Finite Size Behavior». En: *Physical Review Letters* 106.22 (mayo de 2011). DOI: 10.1103/physrevlett.106.225701.

[Gra73] Mark S. Granovetter. «The Strength of Weak Ties». En: *American Journal of Sociology* 78.6 (mayo de 1973), págs. 1360-1380. DOI: 10.1086/225469.

[HDF14] Darko Hric, Richard K. Darst y Santo Fortunato. «Community detection in networks: Structural communities versus ground truth». En: *Physical Review E* 90.6 (2014), págs. 1-19. ISSN: 15502376. DOI: 10.1103/PhysRevE.90.062805.

[HH99] H. P. Hsu y M. C. Huang. «Percolation thresholds, critical exponents, and scaling functions on planar random lattices and their duals». En: *Physical Review E* 60.6 A (1999), págs. 6361-6370. ISSN: 1063651X. DOI: 10.1103/physreve.60.6361.

[HLL83] Paul W. Holland, Kathryn Blackmond Laskey y Samuel Leinhardt. «Stochastic blockmodels: First steps». En: *Social Networks* 5.2 (jun. de 1983), págs. 109-137. DOI: 10.1016/0378-8733(83)90021-7.

[Hol+02] Petter Holme y col. «Attack vulnerability of complex networks». En: *Physical Review E* 65.5 (mayo de 2002). DOI: 10.1103/physreve.65.056109.

[HPF16] Darko Hric, Tiago P. Peixoto y Santo Fortunato. «Network Structure, Metadata, and the Prediction of Missing Nodes and Annotations». En: *Physical Review X* 6.3 (sep. de 2016), pág. 031038. DOI: 10.1103/physrevx.6.031038.

[Iye+13] Swami Iyer y col. «Attack Robustness and Centrality of Complex Networks». En: *PLoS ONE* 8.4 (abr. de 2013). Ed. por Satoru Hayasaka, e59613. DOI: 10.1371/journal.pone.0059613.

[JV95] Wolfhard Janke y Ramon Villanova. «Two-dimensional eight-state Potts model on random lattices: A Monte Carlo study». En: *Physics Letters A* 209.3-4 (dic. de 1995), págs. 179-183. ISSN: 03759601. DOI: 10.1016/0375-9601(95)00813-9.

[JW04] Wolfhard Janke y Martin Weigel. «Harris-Luck criterion for random lattices». En: *Physical Review B - Condensed Matter and Materials Physics* 69.14 (2004), págs. 1-12. ISSN: 01631829. DOI: 10.1103/PhysRevB.69.144208.

[KBD19] Alexander P. Kartun-Giles, Marc Barthelemy y Carl P. Dettmann. «Shape of shortest paths in random spatial networks». En: *Physical Review E* 100.3 (sep. de 2019). DOI: 10.1103/physreve.100.032315.

[Kir+18] Alec Kirkley y col. «From the betweenness centrality in street networks to structural invariants in random planar graphs». En: *Nature Communications* 9.1 (jun. de 2018). DOI: 10.1038/s41467-018-04978-z.

[KKG20] J.-H. Kim, S.-J. Kim y K.-I. Goh. «Critical behaviors of high-degree adaptive and collective-influence percolation». En: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 30.7 (jul. de 2020), pág. 073131. DOI: 10.1063/1.5139454.

[Kor+18] Yosef Kornbluth y col. «Network overload due to massive attacks». En: *Physical Review E* 97.5 (mayo de 2018). DOI: 10.1103/physreve.97.052309.

[Leo+16] Yannick Leo y col. «Socioeconomic correlations and stratification in social-communication networks». En: *Journal of The Royal Society Interface* 13.125 (dic. de 2016), pág. 20160598. ISSN: 1742-5689. DOI: 10.1098/rsif.2016.0598.

[Leo+17] María Juliana Leone y col. «Time to decide: Diurnal variations on the speed and quality of human decisions». En: *Cognition* 158 (2017), págs. 44-55. ISSN: 0010-0277. DOI: http://dx.doi.org/10.1016/j.cognition.2016.10.007.

[Lim+00] F. W.S. Lima y col. «Critical behavior of a three-state Potts model on a Voronoi lattice». En: *European Physical Journal B* 17.1 (2000), págs. 111-114. ISSN: 14346028. DOI: 10.1007/s100510070165.

[LÖ12] Jiantong Li y Mikael Östling. «Corrected finite-size scaling in percolation». En: *Physical Review E* 86.4 (oct. de 2012). DOI: 10.1103/physreve.86.040105.

[LZ21] Ming-Xia Li y Wei-Xing Zhou. «Anatomizing the Elo transfer network of Weiqi players». En: *The European Physical Journal B* 94.8 (ago. de 2021). DOI: 10.1140/epjb/s10051-021-00180-1.

[MAA02] André Auto Moreira, José S. Andrade y Luís A. Nunes Amaral. «Extremum Statistics in Scale-Free Network Models». En: *Physical Review Letters* 89.26 (dic. de 2002). DOI: 10.1103/physrevlett.89.268703.

[Mas22] Opening Master. https://www.openingmaster.com/. [Accedido el 28 de Marzo de 20222]. 2022.

[McC87] J. F. McCarthy. «Invasion percolation on a random lattice». En: *Journal of Physics A: General Physics* 20.11 (1987), págs. 3465-3469. ISSN: 03054470. DOI: 10.1088/0305-4470/20/11/047.

[Meg+01] S. Meguerdichian y col. «Coverage problems in wireless ad-hoc sensor networks». En: *Proceedings IEEE INFOCOM 2001. Conference on Computer Communications. Twentieth Annual Joint Conference of the IEEE Computer and Communications Society (Cat. No.01CH37213)*. Vol. 3. 4. IEEE, 2001, págs. 1380-1387. ISBN: 0-7803-7016-3. DOI: 10.1109/INFCOM.2001.916633.

[Mel13] O. Melchert. «Percolation thresholds on planar Euclidean relative neighborhood graphs». En: *Physical Review E* 87.4 (2013), págs. 1-7. ISSN: 15393755. DOI: 10.1103/PhysRevE.87.042106.

[Mil+03] R. Milo y col. «On the uniform generation of random graphs with prescribed degree sequences». En: *arXiv preprint arXiv:cond-mat/0312028* (2003).

[MM15] Flaviano Morone y Hernán A. Makse. «Influence maximization in complex networks through optimal percolation». En: *Nature* 524.7563 (jul. de 2015), págs. 65-68. DOI: 10.1038/nature14604.

[MN00] C. Moore y M. E. J. Newman. «Epidemics and percolation in small-world networks». En: *Physical Review E* 61.5 (mayo de 2000), págs. 5678-5682. DOI: 10.1103/physreve.61.5678.

[Mor+17] Flaviano Morone y col. «Model of brain activation predicts the neural collective influence map of the brain». En: *Proceedings of the National Academy of Sciences* 114.15 (mar. de 2017), págs. 3849-3854. DOI: 10.1073/pnas.1620808114.

[MR95] M. Molloy y B. Reed. «A critical point for random graphs with a given degree sequence». En: *Proceedings of the Sixth International Seminar on Random Graphs and Probabilistic Methods in Combinatorics and Computer Science, "Random Graphs '93" (Poznań, 1993)*. Vol. 6. 1995, págs. 161-179. DOI: 10.1002/rsa.3240060204.

[NC16] M. E. J. Newman y Aaron Clauset. «Structure and inference in annotated networks». En: *Nature Communications* 7.May (2016), págs. 1-16. ISSN: 2041-1723. DOI: 10.1038/ncomms11863.

[New01] M. E. J. Newman. «The structure of scientific collaboration networks». En: *Proceedings of the National Academy of Sciences* 98.2 (ene. de 2001), págs. 404-409. DOI: 10.1073/pnas.98.2.404.

[New02] M. E. J. Newman. «Assortative Mixing in Networks». En: *Phys. Rev. Lett.* 89.20 (oct. de 2002), pág. 208701. DOI: 10.1103/physrevlett.89.208701.

[New03a] M. E. J. Newman. «Mixing patterns in networks». En: *Physical Review E* 67.2 (feb. de 2003), pág. 026126. DOI: 10.1103/physreve.67.026126.

[New03b] M. E. J. Newman. «The Structure and Function of Complex Networks». En: *SIAM Review* 45.2 (2003), págs. 167-256. DOI: 10.1137/s003614450342480.

[New05] M. E. J. Newman. «Power laws, Pareto distributions and Zipf's law». En: *Contemporary Physics* 46.5 (2005), págs. 323-351. DOI: 10.1080/00107510500052444.

[New18] Mark Newman. *Networks*. 2.ª ed. London, England: Oxford University Press, jul. de 2018.

[NG04] M. E. J. Newman y M. Girvan. «Finding and evaluating community structure in networks». En: *Physical Review E* 69.2 (feb. de 2004). DOI: 10.1103/physreve.69.026113.

[NMH16] C. Norrenbrock, O. Melchert y A. K. Hartmann. «Fragmentation properties of two-dimensional proximity graphs considering random failures and targeted attacks». En: *Physical Review E* 94.6 (2016), págs. 1-11. ISSN: 24700053. DOI: 10.1103/PhysRevE.94.062125.

[Nor16] Christoph Norrenbrock. «Percolation threshold on planar Euclidean Gabriel graphs». En: *European Physical Journal B* 89.5 (2016). ISSN: 14346036. DOI: 10.1140/epjb/e2016-60728-0.

[NZ00] M. E. J. Newman y R. M. Ziff. «Efficient Monte Carlo Algorithm and High-Precision Results for Percolation». En: *Physical Review Letters* 85.19 (nov. de 2000), págs. 4104-4107. DOI: 10.1103/physrevlett.85.4104.

[NZ01] M. E. J. Newman y R. M. Ziff. «Fast Monte Carlo algorithm for site or bond percolation». En: *Physical Review E* 64.1 (jun. de 2001). DOI: 10.1103/physreve.64.016706.

[Oli+08] Marcelo M. de Oliveira y col. «Contact process on a Voronoi triangulation». En: *Physical Review E* 78.3 (sep. de 2008), pág. 031133. ISSN: 1539-3755. DOI: 10.1103/PhysRevE.78.031133.

[Ors+15] Chiara Orsini y col. «Quantifying randomness in real networks». En: *Nature Communications* 6.1 (oct. de 2015). DOI: 10.1038/ncomms9627.

[PAS20] Juan Ignacio Perotti, Nahuel Almeira y Fabio Saracco. «Towards a generalization of information theory for hierarchical partitions». En: *Physical Review E* 101.6 (jun. de 2020). DOI: 10.1103/physreve.101.062148.

[PB12] Tiago P. Peixoto y Stefan Bornholdt. «Evolution of Robust Network Topologies: Emergence of Central Backbones». En: *Phys. Rev. Lett.* 109 (11 2012), pág. 118703. DOI: 10.1103/PhysRevLett.109.118703.

[PC12] Jürgen Pfeffer y Kathleen M. Carley. «k-Centralities». En: *Proceedings of the 21st international conference companion on World Wide Web - WWW '12 Companion*. ACM Press, 2012. DOI: 10.1145/2187980.2188239.

[Per+13] J. I. Perotti y col. «Innovation and nested preferential growth in chess playing behavior». En: *EPL (Europhysics Letters)* 104.4 (2013), pág. 48005. DOI: 10.1209/0295-5075/104/48005.

[Pro12] Frédéric Prost. «On the Impact of Information Technologies on Society: an Historical Perspective through the Game of Chess». En: *Turing-100*. Ed. por Andrei Voronkov. Vol. 10. EPiC Series. EasyChair, 2012, págs. 268-277.

[PVV01] Romualdo Pastor-Satorras, Alexei Vázquez y Alessandro Vespignani. «Dynamical and Correlation Properties of the Internet». En: *Physical Review Letters* 87.25 (nov. de 2001). DOI: 10.1103/physrevlett.87.258701.

[Rav+02] E. Ravasz y col. «Hierarchical Organization of Modularity in Metabolic Networks». En: *Science* 297.5586 (2002), págs. 1551-1555. ISSN: 00368075. DOI: 10.1126/science.1073374.

[Rav04] Erzsébet Ravasz. «Evolution, Hierarchy and Modular Organization in Complex Networks». Tesis doct. University of Notre Dame, sep. de 2004.

[RB03] Erzsébet Ravasz y Albert-László Barabási. «Hierarchical organization in complex networks». En: *Physical Review E* 67.2 (feb. de 2003), pág. 026112. DOI: 10.1103/physreve.67.026112.

[RCR18] L. S. Ramirez, P. M. Centres y A. J. Ramirez-Pastor. «Standard and inverse bond percolation of straight rigid rods on square lattices». En: *Physical Review E* 97.4 (abr. de 2018). DOI: 10.1103/physreve.97.042113.

[RGG15] B. Requião da Cunha, J. C. González-Avella y S. Gonçalves. «Fast Fragmentation of Networks Using Module-Based Attacks». En: *PLOS ONE* 10.11 (2015), págs. 1-15. DOI: 10.1371/journal.pone.0142824.

[Rib+13] Haroldo V. Ribeiro y col. «Move-by-Move Dynamics of the Advantage in Chess Matches Reveals Population-Level Learning of the Game». En: *PLoS ONE* 8.1 (ene. de 2013), e54165. DOI: 10.1371/journal.pone.0054165.

[SA18] Dietrich Stauffer y Ammon Aharony. *Introduction To Percolation Theory*. Taylor & Francis, dic. de 2018. DOI: 10.1201/9781315274386.

[SB05] M Angeles Serrano y Marián Boguná. «Tuning clustering in random networks with arbitrary degree distributions». En: *Physical Review E* 72.3 (2005), pág. 036133. DOI: 10.1103/PhysRevE.72.036133.

[ŠB11] Lovro Šubelj y Marko Bajec. «Robust Network Community Detection Using Balanced Propagation». En: *Eur. Phys. J. B* 81.3 (2011), págs. 353-362. DOI: 10.1140/epjb/e2011-10979-2.

[Sch+11] Christian M. Schneider y col. «Mitigation of malicious attacks on networks». En: *Proceedings of the National Academy of Sciences* 108.10 (feb. de 2011), págs. 3838-3841. DOI: 10.1073/pnas.1009440108.

[SE64] M. F. Sykes y J. W. Essam. «Exact Critical Percolation Probabilities for Site and Bond Problems in Two Dimensions». En: *Journal of Mathematical Physics* 5.8 (ago. de 1964), págs. 1117-1127. DOI: 10.1063/1.1704215.

[SG17] Giovanni Sala y Fernand Gobet. «Does chess instruction improve mathematical problem-solving ability? Two experimental studies with an active control group». En: *Learning & Behavior* (2017), págs. 1-8.

[Sig+10] Mariano Sigman y col. «Response Time Distributions in Rapid Chess: A Large-Scale Decision Making Experiment». En: *Frontiers in Neuroscience* 4 (oct. de 2010), pág. 1. ISSN: 1662-4548. DOI: 10.3389/fnins.2010.00060.

[SPB14] Ana L. Schaigorodsky, Juan I. Perotti y Orlando V. Billoni. «Memory and long-range correlations in chess games». En: *Physica A: Statistical Mechanics and its Applications* 394.0 (2014), págs. 304-311. ISSN: 0378-4371. DOI: 10.1016/j.physa.2013.09.035.

[SPB16] Ana L. Schaigorodsky, Juan I. Perotti y Orlando V. Billoni. «A Study of Memory Effects in a Chess Database». En: *PLOS ONE* 11.12 (dic. de 2016), págs. 1-18. DOI: 10.1371/journal.pone.0168213.

[Sta87] H.E. Stanley. *Introduction to Phase Transitions and Critical Phenomena*. International series of monographs on physics. Oxford University Press, 1987. ISBN: 9780195053166. URL: https://books.google.com.ar/books?id=C3BzcUxoaNkC.

[TK22] Marco Costalba Tord Romstad y Joona Kiiski. *The Stockfish Engine*. https://stockfishchess.org/. [Accedido el 28 de Marzo de 20222]. 2022.

[TT91] T.M. Cover y Joy A Thomas. *Elements of Information Theory*. en. 99.ª ed. Wiley Series in Telecommunications and Signal Processing. Nashville, TN: John Wiley & Sons, sep. de 1991.

[TWE19] V. A. Traag, L. Waltman y N. J. van Eck. «From Louvain to Leiden: guaranteeing well-connected communities». En: *Scientific Reports* 9.1 (mar. de 2019). DOI: 10.1038/s41598-019-41695-z.

[VBB08] Alessandro Vespignani, Alain Barrat y Marc Barthelemy. *Dynamical processes on complex networks*. en. Cambridge, England: Cambridge University Press, oct. de 2008.

[VEB09] Nguyen Xuan Vinh, Julien Epps y James Bailey. «Information theoretic measures for clusterings comparison». En: *Proceedings of the 26th Annual International Conference on Machine Learning - ICML '09*. ACM Press, 2009. DOI: 10.1145/1553374.1553511.

[VL15] Fabien Viger y Matthieu Latapy. «Efficient and simple generation of random simple connected graphs with prescribed degree sequence». En: *Journal of Complex Networks* 4.1 (jun. de 2015), págs. 15-37. DOI: 10.1093/comnet/cnv013.

[VPV02] Alexei Vázquez, Romualdo Pastor-Satorras y Alessandro Vespignani. «Large-scale topological and dynamical properties of the Internet». En: *Phys. Rev. E* 65 (6 jun. de 2002), pág. 066130. DOI: 10.1103/PhysRevE.65.066130.

[Wan+18] Sebastian Wandelt y col. «A comparative analysis of approaches to network-dismantling». En: *Scientific Reports* 8.1 (sep. de 2018). DOI: 10.1038/s41598-018-31902-8.

[Wil05] Herbert S Wilf. *generatingfunctionology*. 3.ª ed. Natick, MA: A K Peters, dic. de 2005.

[WS98] Duncan J. Watts y Steven H. Strogatz. «Collective dynamics of "small-world" networks». En: *Nature* 393.6684 (jun. de 1998), págs. 440-442. DOI: 10.1038/30918.

[XFH10] Y. Xia, J. Fan y D. Hill. «Cascading failure in Watts-Strogatz small-world networks». En: *Physica A* 389.6 (2010), págs. 1281-1285. ISSN: 0378-4371. DOI: 10.1016/j.physa.2009.11.037.

[Yan+15] Yang Yang y col. «Improving the robustness of complex networks with preserving community structure». En: *PLOS ONE* 10.2 (2015), págs. 1-14. DOI: 10.1371/journal.pone.0116551.

[Yoo+01] S. H. Yook y col. «Weighted Evolving Networks». En: *Physical Review Letters* 86.25 (jun. de 2001), págs. 5835-5838. DOI: 10.1103/physrevlett.86.5835.

[Yua+16] Xin Yuan y col. «k-core percolation on complex networks: Comparing random, localized, and targeted attacks». En: *Phys. Rev. E* 93 (6 2016), pág. 062302. DOI: 10.1103/PhysRevE.93.062302.

[Zac77] W. W. Zachary. «An information flow model for conflict and fission in small groups». En: *Journal of anthropological research* 33.4 (1977), págs. 452-473.

[Zen+20] Guanwen Zeng y col. «Multiple metastable network states in urban traffic». En: *Proceedings of the National Academy of Sciences* 117.30 (jul. de 2020), pág. 201907493. ISSN: 0027-8424. DOI: 10.1073/pnas.1907493117.

[ZL12] An Zeng y Weiping Liu. «Enhancing network robustness against malicious attacks». En: *Phys. Rev. E* 85 (6 2012), pág. 066130. DOI: 10.1103/PhysRevE.85.066130.

[ZM04] S. Zhou y R.J. Mondragon. «The Rich-Club Phenomenon in the Internet Topology». En: *IEEE Communications Letters* 8.3 (mar. de 2004), págs. 180-182. DOI: 10.1109/lcomm.2004.823426.