

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

COMUNIDADES EN GRAFOS

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL MATEMÁTICO

RODOLFO ANGHELLO CARVAJAL VIEYTE

PROFESOR GUÍA:  
IVÁN RAPAPORT ZIMERMANN

MIEMBROS DE LA COMISIÓN:  
MARTÍN MATAMALA VÁSQUEZ  
MARCOS KIWI KRAUSKOPF

SANTIAGO DE CHILE  
OCTUBRE 2006

# Agradecimientos

Probablemente sea difícil nombrar a todos aquellos a los que debo mi gratitud, por haber ayudado de alguna u otra forma a que hoy me encuentre al final de esta etapa, que en cierta manera comenzó el día en que nací. Trataré de aproximarme lo mejor que pueda.

Agradezco a Iván Rapaport, por ser un apoyo constante durante el desarrollo de mi memoria, permitiéndome tener la tranquilidad y libertad necesarias para trabajar, además por su paciencia y disposición para con mi inexperiencia e ineficiencia. Agradezco también a Martín Matamala y Marcos Kiwi, primero por despertar en mí el gusto por las matemáticas discretas y segundo, por estar siempre dispuestos a ayudarme tanto durante mi memoria, como durante toda mi carrera.

Así también, agradezco a todos los docentes del departamento, por ser fuente de inspiración y admiración en sus propias áreas, en especial a aquellos de los cuales tuve el honor de ser alumno. No puedo dejar de agradecer además a nuestros funcionarios, quienes hacen posible que el departamento ande día a día.

Toda mi gratitud también a mis amigos, que a lo largo de toda mi vida han estado ahí. Primero mis nobles y viejos amigos de la etapa del colegio, luego mis queridos amigos de plan común y en el último tiempo mis brillantes amigos de la carrera, a quienes admiro y respeto profundamente.

Por supuesto debo agradecer a mi adorada familia, sobre todo por su paciencia para con mi memoria. A mis hermanos Ashley y Yordi Paulo, por ser siempre mis mejores amigos. A cada momento recuerdo y recordaré todas las aventuras que hemos pasado juntos. A mis padres Bernarda y Víctor, mis verdaderos héroes, por dármelo todo siempre e incondicionalmente (incluso aquellos valiosos consejos que, torpemente, recibo con rebeldía e incluso ignoro). Espero llegar a formar una familia tan hermosa como la que ustedes han logrado formar. Siempre estarán en mi corazón. Finalmente a mi amada novia (y por ende futura esposa) Pamela, por todo lo que haces por mí, desde apoyarme, levantarme cuando caigo y ordenar mi vida, hasta hacerme feliz con tu mirada y tu sonrisa o con el solo hecho de saber que tu corazón irriga sangre a tu cerebro. Eres la mujer con la que quiero caminar por la vida, para siempre.

Me permito también agradecer a los nobles y anónimos desarrolladores de software libre, quienes me han permitido gozar de un entorno estable y confiable de trabajo.

Por último, agradezco al proyecto Anillo: Redes en Matemáticas y Ciencias de la Ingeniería ACT08, por financiar esta memoria.

*Phil: No matter what happens tomorrow or for the rest of my life,  
I'm happy now.*

*Bill Murray como Phil Connors en Groundhog Day*

Este trabajo está dedicado a mis abuelos, que son los que han muerto,  
a mis padres, a mis hermanos y a mi futura esposa.

# Índice general

<b>1. Definiciones y Notaciones Preliminares</b>	<b>1</b>
<b>2. Introducción</b>	<b>2</b>
2.1. El coeficiente de clustering . . . . .	3
2.2. El voto de mayoría y comunidades . . . . .	7
2.3. Detección de comunidades y partición de grafos . . . . .	8
<b>3. Comunidades en Grafos</b>	<b>11</b>
3.1. Grafos con grado alto . . . . .	12
3.2. Grafos planares regulares de grado 3 y 4 . . . . .	14
3.3. Árboles . . . . .	16
3.4. Conjeturas . . . . .	18
<b>4. El Coeficiente de Comunidad</b>	<b>19</b>
4.1. Un primer acercamiento . . . . .	19
4.2. Definiendo $cc(G)$ . . . . .	21
4.3. Calculando $cc(G)$ . . . . .	22
<b>5. Resultados Experimentales</b>	<b>25</b>

## ÍNDICE GENERAL

---

5.1. $\alpha$ -Grafos . . . . .	25
5.2. $\beta$ -Grafos . . . . .	27
<b>6. Grafos Planares sin Puentes Regulares de Grado 3 y 4</b>	<b>29</b>
<b>Bibliografía</b>	<b>36</b>

# Capítulo 1

## Definiciones y Notaciones Preliminares

El presente trabajo asume un cierto conocimiento en teoría de grafos y complejidad computacional, sin embargo es saludable indicar cuáles son las definiciones y notaciones que se usarán. Para más detalles o nociones básicas acerca de teoría de grafos, consultar [1]. Para complejidad computacional, consultar [2].

Sea un grafo  $G = (V, E)$ , con  $V$  el conjunto de vértices y  $E$  el de aristas. En general denotamos  $V(G) = V$  y  $E(G) = E$ . Dado un subconjunto de vértices  $S \subseteq V$ ,  $G(S)$  representa el subgrafo de  $G$  inducido por los vértices de  $S$ . Denotamos al *corte* inducido por  $S$  como  $E(S, \overline{S}) = \{(u, v) \in E : u \in S \wedge v \in \overline{S}\}$ .

Dado un vértice  $v \in V$ , el grafo  $G \setminus v$  se define como el grafo que se obtiene de remover al vértice  $v$  de  $G$  y todas sus aristas incidentes. Es decir,  $G \setminus v := (V \setminus \{v\}, E \setminus \{(v, u) \in E : u \in V\})$ .

La *vecindad* de  $v$  en  $G$ , será denotada por  $N_G(v) := \{u \in V(G) : (u, v) \in E(G)\}$ . El *grado* de  $v$  en  $G$ , se denotará  $d_G(v) = |N_G(v)|$ . Esta notación se extiende naturalmente para un conjunto de vértices  $S \subseteq V$ :  $d_S(v) = |N_G(v) \cap S|$ . Un grafo se dice *regular* si todos sus vértices tienen igual grado  $d$ . Si  $d = 3$ , el grafo se dice *cúbico*.

La *distancia* entre los vértices  $u$  y  $v$  en el grafo  $G$ , se denotará  $d_G(u, v)$ . La *eccentricidad* de  $v$  en  $G$  será denotada  $ecc_G(v) = \max_{u \in V(G)} d_G(u, v)$ . Esta última notación también se extiende naturalmente a  $S \subseteq V$  como  $ecc_S(v) = ecc_{G(S)}(v)$ . El *radio* de  $G$ , se denotará  $r(G) = \min_{v \in V(G)} ecc_G(v)$  y el *diámetro* de  $G$ , será denotado  $diam(G) = \max_{v \in V(G)} ecc_G(v)$ .

Para cualquier otra definición que haya sido omitida aquí, se sugiere revisar las referencias señaladas al principio de este capítulo.

# Capítulo 2

## Introducción

El uso de grafos en la modelación de diversos problemas y fenómenos, es natural y cada vez más frecuente en distintas áreas. Algunos ejemplos son Internet, la *World Wide Web*, redes sociales entre individuos, redes neuronales, redes metabólicas, redes de alimentación, redes organizacionales y de negocios, redes de distribución, redes de citas entre artículos científicos, entre otros. En dichos modelos, los vértices del grafo podrían representar a los objetos o individuos involucrados y las aristas a las relaciones entre estos. Para excelentes revisiones acerca del tema y referencias, ver [3, 4].

El estudio de redes complejas, tanto de su estructura como función, ha cambiado considerablemente en el último tiempo por diversas razones. Una primera razón es la diferencia entre las estructuras topológicas usadas. En los inicios, se asumía que las relaciones de interconexión entre individuos podían ser representadas en estructuras regulares como lattices e hipercubos, entre otros. No fue hasta los años 50s que como forma de modelar las interacciones, se adopta el uso de grafos aleatorios (herramienta matemática debida a Erdős y Rényi [5]). Sin embargo, en los últimos años, se descubre que gran parte de las redes encontradas en fenómenos reales manifiestan un comportamiento intermedio, entre el orden y la aleatoriedad [6, 7, 8]. Estas conclusiones han sido muchas veces motivados por la posibilidad (debida al aumento del poder de cálculo), de realizar experimentos estadísticos sobre redes reales.

Otra razón del cambio mencionado, es que en lugar de estudiarse propiedades de grafos pequeños y de vértices particulares de ellos, se ha dirigido la atención hacia propiedades estadísticas de grafos a gran escala. Como fue observado antes, esto se ve motivado por la introducción de computadores, que hacen factible el análisis de grandes cantidades de datos. Esto trae consigo el hecho de que preguntas que antes parecían interesantes en el estudio de redes, hoy sólo tienen sentido si son planteadas con un enfoque estadístico. Un ejemplo (mencionado en [3]) es el de la pregunta: ¿Qué vértice es más crucial para la conectividad del

grafo, si es removido?, que en el contexto actual tiene más sentido como: ¿Qué porcentaje de vértices es necesario remover para sustancialmente afectar la conectividad del grafo de alguna manera dada?

Un tercer aspecto es que, pese a que la inspección visual es de gran ayuda para analizar redes de cierto tamaño razonable, redes de millones o billones de vértices son imposibles de examinar a ojo desnudo. Incluso al tratar de usar las potencialidades de la generación de imágenes vía computadores, los dibujos obtenidos muy poco dirán acerca de la estructura de la red. Es decir, vía una aproximación estadística, se intenta saber cómo se ve una red compleja sin poder mirarla.

Como se puede apreciar, el común denominador es la introducción de técnicas estadísticas al estudio de redes complejas. Sin embargo, es preciso señalar que este enfoque estadístico tiene limitaciones y no permite estudiar todas las propiedades importantes que se buscan. Un aspecto particular que nos interesa en este trabajo, es el de obtener medidas de propiedades estructurales del grafo que permitan predecir su comportamiento global, ante reglas que involucran localmente a sus vértices. En otras palabras, el identificar estructura local en este.

Una medida emblemática de estructura local en grafos es el llamado *coeficiente de clustering*, la cual pasamos a describir.

## 2.1. El coeficiente de clustering

Dado un grafo  $G = (V, E)$  con  $n$  vértices, el *coeficiente de clustering* de un vértice  $v$  de grado  $d$  (denotado  $c(v)$  e introducido en [7]), es definido como el cociente entre el número de aristas  $e_v$  entre vecinos de  $v$  y su máximo valor posible  $\frac{d(d-1)}{2}$ . Es decir:

$$c(v) = \frac{2e_v}{d(d-1)}.$$

El promedio de  $c(v)$  sobre todos los vértices del grafo se define como el coeficiente de clustering de  $G$ ,  $c(G)$ . Este número, indica en qué medida los vecinos de un vértice  $v$ , son vecinos entre sí. Esto, en términos de redes sociales, indica qué tan probable es que los amigos de un individuo sean también amigos. Además, el coeficiente da cierta información acerca de la estructura del grafo. De hecho, Watts y Strogatz lo usaron en [7] para estudiar empíricamente el fenómeno del *pequeño mundo*. Este fenómeno dice relación con grafos que son localmente conectados a diferencia de grafos aleatorios clásicos (y que en consecuencia tienen un coeficiente de clustering más alto que estos últimos), mientras que sin embargo ambos tipos de grafos presentan un diámetro pequeño.

Numerosas han sido las menciones en la literatura respecto de las limitaciones que presenta el coeficiente y a la vez propuestas para generalizar el mismo. En [9], se introduce una

definición que intenta filtrar el efecto de la correlación con el grado de los vértices. Es decir, el coeficiente tradicional para vértices de grado alto conectados con vértices de grado mucho menor será siempre pequeño, independiente de la manera en que estén conectados sus vecinos. Un ejemplo de este hecho se presenta en la Figura 2.1, en donde el coeficiente de clustering del vértice 1 es  $c(1) = O(1/n)$  y por lo tanto, decreciente a cero con  $n \rightarrow \infty$ . Sin embargo, los vecinos de 1 están todo lo conectados que sus grados permiten.

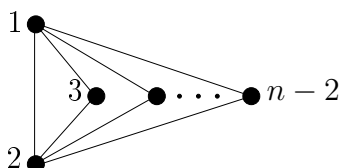


Figura 2.1: Doble estrella con dos vértices.

La definición de este nuevo coeficiente ( $\tilde{c}(v)$ ) se basa en considerar (para un vértice  $v$ ) en lugar del número total de posibles aristas entre sus vértices vecinos, este mismo número pero condicionado al grado de los vecinos de  $v$ . Es decir, si llamamos a este último número  $w_v$ ,

$$\tilde{c}(v) = \frac{e_v}{w_v}.$$

Notemos que un vecino  $u$  de  $v$ , puede tener a lo más  $\min(d_G(v) - 1, d_G(u) - 1)$  aristas hacia otros vecinos de  $v$ . Luego

$$w_v \leq \left\lfloor \frac{1}{2} \sum_{u \in N_G(v)} [\min(d_G(v), d_G(u)) - 1] \right\rfloor \leq \binom{d_G(v)}{2}.$$

Es decir,  $w_v$  considera no solo el grado de  $v$ , sino que no todas las aristas en vecinos de  $v$  están disponibles y que además las aristas disponibles deben formar triángulos. Por otra parte, en [10] se introduce un *coeficiente de clustering con pesos* ( $c_w(v)$ ), para una función  $w : V \rightarrow \mathbb{R}_+$  sobre los vértices. Este coeficiente, tiene la misma definición local que el coeficiente tradicional, la diferencia se encuentra al momento de promediar sobre todos los vértices del grafo:

$$c_w(v) = \frac{1}{\sum_{v \in V} w(v)} \sum_{v \in V} w(v) c(v)$$

Una posible función de peso de la cual sólo hace mención en [10] es la del grado, es decir  $w(v) = d_G(v)$ . Estos pesos, sugieren una forma de tratar el problema de la correlación de  $c(v)$  con el grado de los vértices.

Otra limitación fundamental del coeficiente de clustering es expuesta en la Figura 2.2. En este grafo se observa una clara estructura local ( $k$  está fijo), sin embargo como sus ciclos

más pequeños son de largo  $k + 1$ , si  $k > 2$ , su coeficiente de clustering es 0. Lo mismo sucede en grafos con estructura de lattice, con cintura grande. Para atacar este problema, se ha propuesto cambiar la pregunta que motiva la definición del coeficiente de clustering. En lugar de preguntarse: “¿Cuántos de mis vecinos están conectados?” otros autores en [11] comenzaron a preguntarse: “¿Qué tan cercanamente relacionados se encuentran mis vecinos?”.

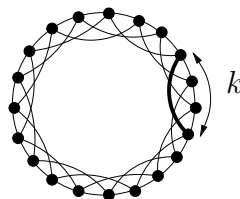


Figura 2.2: Un grafo con estructura local.

Con esta motivación detrás, en [12] se introduce el llamado *coeficiente de cuadrícula*, denotado para un vértice  $v$  como  $c_4(v)$ . A diferencia del coeficiente de clustering que cuenta cuántos ciclos de largo 3 hay en el grafo que contengan a  $v$  (esta es precisamente la cantidad  $e_v$ ), el coeficiente de cuadrícula cuenta cuántos ciclos de largo 4 hay que cumplan la misma condición (dividiendo en este caso por el número de ciclos de largo 4 posibles que contienen a  $v$ ). Como se supone, sin embargo el coeficiente de cuadrícula no logra resolver el problema en casos más generales (como el de la Figura 2.2, con  $k$  mucho mayor).

Una generalización natural (aunque levemente distinta) es la introducida en [13], que consiste en estudiar *coeficientes de clustering de orden mayor*. Un coeficiente de orden  $i \geq 3$  para un vértice  $v$ , es calculado contando cuántos pares de vértices en  $N_G(v)$  tienen una distancia  $i$  en el grafo  $G \setminus v$  (es decir borrando  $v$  y todas sus aristas incidentes) y normalizando por el número total de pares. Es decir:

$$c(v, i) = \frac{|\{(u, w) : u, w \in N_G(v), d_{G \setminus v}(u, w) = i\}|}{\binom{d_G(v)}{2}}$$

Los autores sólo aplican estos coeficientes a grafos de Barabási-Albert, modelo de grafo aleatorio que comienza a crecer desde un pequeño grafo completo de tamaño  $m$  y agregando vértices nuevos que se conectan con  $m_c \leq m$  vértices ya antes agregados, con un esquema de conexión preferencial, es decir que la probabilidad de conectarse a un vértice es proporcional al grado de este (ver [14, 15]). Se concluye que el grafo no presenta un coeficiente de clustering (o coeficiente de orden 1) decreciente a 0, con el tamaño del grafo. Sin embargo, se observa que coeficientes de orden mayor crecen con el tamaño del grafo.

En [11], la misma definición es usada y se llama a los coeficientes de orden superior *perfil de clustering*. Mediante resultados experimentales del cálculo de dicho perfil para ciertos grafos reales, los autores exhiben evidencias de que en estos grafos cuyo coeficiente de clustering es

cercano a 0, también la estructura local se ha desplazado hacia coeficientes  $c(v, i)$  de orden mayor.

Otras extensiones del coeficiente de clustering son debidas a estudios de fenómenos particulares. En [16], se introduce una medida de clustering para grafos planares, motivada por el estudio de la estructura de galerías construidas por hormigas. El coeficiente, denotado  $M$  es el cociente entre el número de caras en el grafo (sin contar la exterior)  $f$  y el número máximo de caras que puede tener el grafo  $f_{max} = 2n - 5$ , es decir:

$$M = \frac{f}{f_{max}}.$$

En [17], más que intentar generalizar el coeficiente de clustering se busca una forma de estudiar el fenómeno del “pequeño mundo”, vía una medida única. Como vimos anteriormente, este fenómeno dice relación con dos aspectos: un coeficiente de clustering alto y diámetro pequeño. El enfoque usado es el de la *eficiencia*, en términos de qué tan eficiente es el grafo para intercambiar información. Para cada par de vértices  $(u, v)$ , se considera la eficiencia de comunicación entre  $u$  y  $v$  como el inverso de la distancia entre estos en el grafo ( $\epsilon_{u,v} = \frac{1}{d_G(u,v)}$ ). En el caso en que  $d_G(u, v) = +\infty$ , se define de manera consistente  $\epsilon_{uv} = 0$ . Luego, la eficiencia promedio (o eficiencia global) del grafo es:

$$E(G) = \frac{2}{n(n-1)} \sum_{u \neq v \in V} \epsilon_{uv}.$$

Además, para el caso de grafos no conexos, se propone considerar la eficiencia local de un vértice como  $E(G(N(v)))$  (es decir, la eficiencia del grafo inducido por los vecinos de  $v$ ) y luego, la eficiencia local del grafo,  $E_{loc}(G)$ , como el promedio sobre todos los vértices. De esta manera, los autores reformulan la noción de grafo tipo “pequeño mundo” como aquellos grafos que tienen ambas eficiencias, global y local grandes. Es decir, grafos que son eficientes comunicando información global y localmente. Se presentan además resultados experimentales, calculando las eficiencias para un modelo de grafo aleatorio que mediante un parámetro ajustable, manifiesta el fenómeno del pequeño mundo (usado en [7] y que estudiamos en el Capítulo 5). Se observa en dichos resultados indicios de que la eficiencia permite detectar este último fenómeno.

Nuestro enfoque en el presente trabajo es diferente y se basa en el concepto de *comunidad*, definido en [18] como:

*Una comunidad es un conjunto de vértices que tienen más vínculos hacia miembros de la comunidad que a no-miembros.*

Este concepto motiva la siguiente observación: los vértices de grafos localmente estructurados pertenecen a *pequeñas comunidades*.

En lo que sigue, estudiamos propiedades respecto del tamaño de dichas comunidades en grafos y de qué tan probable es encontrar una comunidad en un conjunto de vértices escogido aleatoriamente entre los del grafo. Luego, una primera medida de estructura local de un vértice  $v$  en un grafo  $G$  que se sugiere, es el tamaño de la comunidad más pequeña que contiene a  $v$ . Sin embargo ante la imposibilidad de calcular eficientemente dicho número, se propone en cambio calcular qué tan lejos de  $v$  se puede encontrar una comunidad que lo contenga. Formalmente, el menor  $k \in \mathbb{N}$  tal que el conjunto formado por todos los vértices a distancia menor o igual que  $k$  de  $v$ , contiene una comunidad a la que pertenezca  $v$ . Denotamos a dicho valor  $depth_G(v)$ . Este número resulta ser calculable a tiempo polinomial, mediante un algoritmo que aquí se exhibe, lo que motiva definir una medida de estructura local tomando  $depth_G(v)$  normalizada y promediándola sobre los vértices del grafo. A dicha medida la llamamos *coeficiente de comunidad*. Finalmente se estudia dicho coeficiente tanto experimental como analíticamente, para ciertas clases de grafos. Dicha noción de comunidad parece bastante natural para situaciones en las que el concepto de localidad está involucrado y no es sorprendente encontrarla en otros contextos. A continuación se presentan algunos de ellos (o los dos más importantes).

## 2.2. El voto de mayoría y comunidades

El proceso de voto de mayoría en grafos es, dado un grafo  $G = (V, E)$  en el que sus vértices se pueden encontrar en dos posibles estados: 0 ó 1, la aplicación iterativa de una regla de mayoría para la actualización de estados. El estado siguiente de un vértice estará dado por el estado que sea mayoría entre sus vecinos (incluido él mismo). Se distingue entre voto de mayoría *simple* y *fuerte*, para describir si un vértice en estado 0 cambia o no a estado 1, en caso de haber igual número de vecinos en ambos estados. Además, se habla de voto *reversible* e *irreversible* si se permite o no que vértices en estado 1 cambien a estado 0. Esta dinámica tiene particular relación con algoritmos falla-tolerantes y cálculo distribuido. Una pregunta de interés es cuál es la posibilidad que tiene un conjunto de vértices “con fallas” de corromper a la mayoría de los vértices cercanos. Así, se introducen las nociones de coalición y monopolio en grafos, definidas de la siguiente manera [19, 20]:

Dado un vértice  $v \in V$  y  $M \subseteq V$ , se dice que  $M$  es *coalición que controla a  $v$* , si al menos la mitad de los vecinos de  $v$  se encuentran en  $M$ . Si un conjunto  $M$  es coalición que controla a todo vértice de  $G$ , se dice *monopolio*.

Una extensión dinámica de coalición y monopolio es el de conjuntos de vértice que controlan el grafo pero con un retardo en las actualizaciones de estados. Es decir, un conjunto  $M \subseteq V$  tal que si todos sus vértices son puestos inicialmente en estado 1 (y el resto en estado 0), al cabo de un número finito de aplicaciones de la regla de mayoría, llevan a todos los vértices del grafo a un estado fijo 1. Dichos conjuntos se llaman *monopolios dinámicos* o

*dínamos* [21]. En la Figura 2.3 se presenta un ejemplo: el conjunto  $M = \{v_1, v_2, v_3\}$  es un dínamo. Es aquí en donde la idea de comunidad se manifiesta. Basta notar que (algo que trataremos en detalle más adelante, en el Capítulo 4) si  $S$  es una comunidad en el grafo  $G$  y ponemos a todos sus vértices en estado 0 al inicio, es imposible que estos vértices cambien alguna vez a estado 1 (aplicando la regla de mayoría simple) sin importar cuál sea la distribución de estados de los vértices en  $V \setminus S$ . Esto, ya que cada vértice en  $S$  tiene la mayoría de sus vecinos en la misma comunidad y por ende en estado 0. Entonces, cualquier dínamo en  $G$ , debe necesariamente contener al menos un vértice en  $S$  y de hecho en cualquier comunidad disjunta de  $S$  que exista en  $G$ . Es precisamente este argumento el que se usa en [22] (llamando a las comunidades *bloques blancos*), para encontrar cotas inferiores para el tamaño de dínamos en topologías toroidales. Además, en [23] se usan comunidades pero con el nombre de *grafos inmunes*. Este último trabajo estudia además el trade-off entre tamaño de un dínamo y tiempo (en aplicaciones de la regla de mayoría) que este requiere para llevar el grafo al estado final fijo.

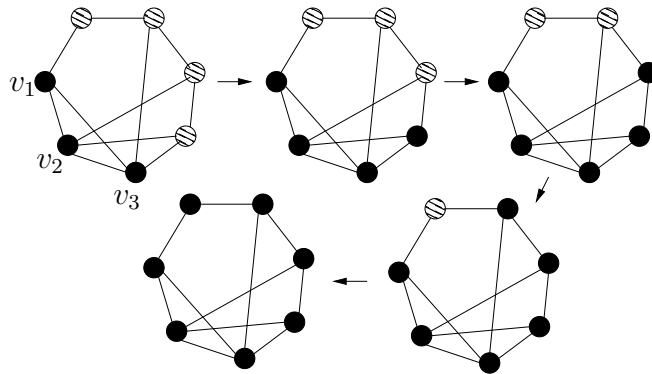


Figura 2.3: Ejemplo de un dínamo. Vértices en negro se encuentran en estado 1, el resto en estado 0.

Un problema interesante en este contexto, a la luz de nuestro trabajo, sea quizás relacionar el comportamiento del coeficiente de comunidad con el tamaño mínimo de un dínamo en el grafo.

Otro ámbito en el que las comunidades juegan un rol central es en el problema de su detección y su uso en particionar grafos.

### 2.3. Detección de comunidades y partición de grafos

La estructura de comunidades en grafos es un tópico particularmente transversal a diversas áreas, esto pues la noción de comunidad representa, según el contexto a ciclos de redes

metabólicas, grupos de páginas web de un mismo tema, grupos de alta actividad en redes de contacto sexual, grupos de libros acerca de temas políticos de tendencia similar, etc. Además la estructura de comunidades de un grafo, genera un meta-grafo (en el que vértices son comunidades) que entrega información gruesa acerca de este. Este aspecto es usado como medio para visualización y análisis de datos [24] .

Es por esto que diversas herramientas matemáticas y algoritmos han sido desarrollados para detectar y dimensionar la estructura de comunidades en grafos. Algunas técnicas comprenden algoritmos de flujo máximo, caminatas aleatorias, optimización, entre otros. Esbozaremos sólo algunas aquí, para más detalles y referencias se recomienda el artículo-revisión de Newman [25] al respecto.

En [18], además de presentarse la definición de comunidad usada en el presente trabajo, se utiliza un método basado en el teorema de *flujo máximo-corte mínimo* de Ford y Fulkerson para detectar comunidades. El método se basa en la siguiente observación:

*Una comunidad,  $S$ , puede ser identificada al calcular el mínimo  $s - t$  corte de  $G$  con  $s$  y  $t$  usados como fuente y sumidero. Después del corte, los vértices que sean alcanzables desde  $s$ , pertenecen a la comunidad.*

(De hecho un enfoque similar es utilizado en el presente trabajo para estudiar el tamaño de comunidades, en el Capítulo 3). Así, el problema de detectar comunidades se reduce a resolver un problema de flujo máximo. Los autores aplican este método para aproximar la detección de comunidades en la *World Wide Web*. Además, en [26, 27], se introducen restricciones a la definición de comunidad para evitar ambigüedades en la frontera de estas.

Una segunda aproximación a la detección de comunidades y partición de grafos es aquella usando un paradigma de optimización. Nuevamente se busca minimizar la cantidad de aristas entre dos grupos de vértices, digamos  $S_1$  y  $S_2$ , para ello se considera la matriz de adyacencia  $A$ :

$$A_{ij} = \begin{cases} 1, & \text{si hay una arista entre } i \text{ y } j, \\ 0, & \text{en otro caso.} \end{cases}$$

Por lo tanto el tamaño del corte entre estos dos grupos de vértices es

$$R = \frac{1}{2} \sum_{i \in S_1, j \in S_2} A_{ij}.$$

Luego, por conveniencia se define un vector índice (que determina a  $S_1$  y  $S_2$ ),  $s$ , dado por:

$$s_i = \begin{cases} 1, & \text{si } i \in S_1, \\ -1, & \text{si } i \in S_2. \end{cases}$$

De esta manera es posible escribir,

$$R = \frac{1}{2}s^t L s,$$

con una cierta matriz  $L$  simétrica y definida positiva, llamada *matriz Laplaciana* del grafo. Así, el problema de minimizar corte, se reduce a encontrar  $s$  vía descomponerlo en una base de vectores propios de  $L$  y poniendo coordenadas más grandes en vectores propios asociados a valores propios más pequeños. Para particionar el grafo en un número mayor de comunidades, se repite este procedimiento dentro de cada una de las comunidades encontradas antes.

En [25], se recurre al mismo enfoque anterior pero con una función objetivo distinta: la *modularidad*. Esta cantidad, denotada por  $Q$ , grosso modo es:

$$Q = (\text{Número de aristas dentro de los grupos}) - (\text{Número esperado de aristas}),$$

en donde el número de aristas esperado dice relación con un grafo aleatorio de referencia. Nuevamente, es posible escribir la modularidad como:

$$Q = \frac{1}{2m}s^t B s,$$

con  $B$ , la llamada *matriz de modularidad*. Esta matriz es simétrica pero, a diferencia de la Laplaciana, no posee todos sus valores propios positivos. Se desarrolla un enfoque espectral tal como antes, para maximizar  $Q$ , pero haciendo algunas modificaciones necesarias. El autor, presenta además experimentos en los que se aplica la maximización de la modularidad, observándose mejoras en la partición de grafos en comunidades.

A modo de relacionar nuestro trabajo con el tópico recién descrito, podemos observar que este enfoque de búsqueda de comunidades está basado en particionar el grafo, o ver el grafo en términos de sus comunidades. Por otra parte, lo que se busca aquí es estudiar la estructura local que subyace en el grafo, permitiendo que comunidades en torno a vértices se intersecten.

Finalmente, esta sucinta revisión de problemáticas y aproximaciones relativas al estudio de la estructura de redes complejas (y en particular el aspecto local de estas), muestra que las preguntas por hacer son aún muchas y la transversalidad del tópico, que trae como consecuencia la posibilidad de mezclar herramientas provenientes de las más diversas áreas, es de gran ayuda en la búsqueda de respuestas.

## Capítulo 3

# Comunidades en Grafos

Dado  $G = (V, E)$  grafo. Se dirá que un subconjunto no vacío  $S \subseteq V$  es una *comunidad* de  $G$ , si satisface que todo vértice en  $S$  tiene al menos tantos vecinos en  $S$ , como en  $\bar{S} = V \setminus S$ . En otras palabras,

$$\forall v \in S, \quad d_S(v) \geq d_{\bar{S}}(v)$$

Una primera pregunta que surge es: ¿cuál es el tamaño de la comunidad más pequeña en  $G$ ? Se observa primero que el grado mínimo del grafo introduce una restricción natural sobre dicho tamaño.

**Proposición 3.1.** *Si  $G$  tiene grado mínimo  $d$ , entonces toda comunidad  $S \subseteq V$  satisface que  $|S| \geq 1 + \lceil \frac{d}{2} \rceil$ .*

*Demostración.* Basta tomar  $v \in S$ . Para que la condición de comunidad se cumpla debe tenerse que  $d_S(v) = |N(v) \cap S| \geq \lceil d_G(v)/2 \rceil$ .

Así,

$$|S| \geq |\{v\} \cup (N(v) \cap S)| \geq 1 + \lceil d(v)/2 \rceil \geq 1 + \left\lceil \frac{d}{2} \right\rceil$$

■

Consideramos a continuación la estrella  $K_{1,n-1}$ , con  $n$  par. Claramente, existe una comunidad en el grafo, de tamaño  $\lceil \frac{n-1}{2} \rceil + 1 = \lfloor \frac{n}{2} \rfloor + 1 = \lceil \frac{n}{2} \rceil + 1$ , basta tomar el vértice central (aquel de grado  $n - 1$ ) y  $\lceil \frac{n-1}{2} \rceil$  de sus vecinos. Sin embargo, no es posible encontrar comunidades más pequeñas. Cualquier comunidad en dicho grafo debe necesariamente contener al vértice central, ya que todos los demás vértices lo tienen como vecino. Luego, usando el razonamiento de la proposición anterior y como el grado de dicho vértice es  $n - 1$ , toda

comunidad tendrá tamaño al menos  $\lceil \frac{n-1}{2} \rceil + 1 = \lceil \frac{n}{2} \rceil + 1$ .

Finalmente se concluye que el tamaño de la comunidad más pequeña en  $K_{1,n-1}$  tiene tamaño  $\lceil \frac{n}{2} \rceil + 1$ . Este es un ejemplo de un grafo con comunidades más pequeñas grandes, de hecho muestra que la siguiente cota superior para el tamaño de una comunidad, es justa.

**Proposición 3.2.** *Todo grafo  $G$ , posee una comunidad  $S \subseteq V$  tal que  $|S| \leq \lceil \frac{n}{2} \rceil + 1$ .*

*Demostración.* Consideremos el conjunto de todos los cortes “casi balanceados” de  $G$ , es decir:

$$\mathcal{C} = \{E(U, \bar{U}) : U \subseteq V, |U| \in \{\lceil n/2 \rceil, 1 + \lceil n/2 \rceil\}\}$$

Sea entonces  $U_0 \subseteq V$  tal que  $E(U_0, \bar{U}_0)$  sea corte mínimo en  $\mathcal{C}$ . Más precisamente:

$$|E(U_0, \bar{U}_0)| = \min_{E \in \mathcal{C}} |E|$$

Afirmamos que  $U_0$  es una comunidad de  $G$ , lo cual demuestra lo requerido.

De no ser así, existe  $u \in U_0$  tal que  $d_{U_0}(u) < d_{\bar{U}_0}(u)$ . Definimos entonces  $U_1 = U_0 \setminus \{u\}$ , cuyo corte asociado satisface:

$$|E(U_1, \bar{U}_1)| = |E(U_0, \bar{U}_0)| - (d_{\bar{U}_0}(u) - d_{U_0}(u)) < |E(U_0, \bar{U}_0)|$$

ya que  $d_{\bar{U}_0}(u) - d_{U_0}(u) > 0$ .

Además, si  $|U_0| = 1 + \lceil n/2 \rceil$  entonces  $|U_1| = \lceil n/2 \rceil$ . Por otra parte si  $|U_0| = \lceil n/2 \rceil$  luego  $|\bar{U}_1| \in \{\lceil n/2 \rceil, 1 + \lceil n/2 \rceil\}$ , por lo que  $E(U_1, \bar{U}_1) \in \mathcal{C}$ , lo cual contradice la minimalidad de  $E(U_0, \bar{U}_0)$ . ■

La comunidad obtenida en la Proposición 3.2, es el resultado de una construcción muy particular. ¿Qué pasa si seleccionamos aleatoriamente un conjunto de vértices grande? ¿Contendrá una comunidad de  $G$  con alta probabilidad? A continuación formalizamos esta pregunta.

Sea  $p \in [0, 1]$ . Denotemos por  $V_p(G)$  al conjunto resultante de seleccionar cada vértice de  $V$  con probabilidad  $p$ . Denotemos también  $r_p(G) = \mathbb{P}\{V_p(G) \text{ contiene una comunidad}\}$ . El problema de calcular  $r_p(G)$  parece ser bastante difícil en general, sin embargo puede ser atacado en algunos casos. Por ejemplo, si el grado de todos los vértices de  $G$  es alto.

### 3.1. Grafos con grado alto

Primero, veamos el siguiente lema (que como nos sugiriese Nicolas Schabanel es un enfoque estándar que aparece, por ejemplo, en [28]):

**Lema 3.3.** Sea  $G = (V, E)$  un grafo de grado mínimo  $d$ . Sea  $\frac{1}{2} < p < 1$ . Luego  $r_p(G) \geq 1 - ne^{-\frac{p\delta^2}{2}d}$ , en donde  $\delta > 0$  es tal que  $\frac{1}{2} = p(1 - \delta)$ .

*Demostración.* Veamos

$$\begin{aligned} r_p(G) &\geq \mathbb{P}\{\forall v \in V_p(G) : d_{V_p(G)}(v) \geq d_{\overline{V_p(G)}}(v)\} \\ &= 1 - \mathbb{P}\{\exists v \in V_p(G) : d_{V_p(G)}(v) < d_{\overline{V_p(G)}}(v)\} \\ &\geq 1 - \sum_{v \in V_p(G)} \mathbb{P}\{d_{V_p(G)}(v) < d_{\overline{V_p(G)}}(v)\} \\ &\geq 1 - \sum_{v \in V_p(G)} \mathbb{P}\left\{d_{V_p(G)}(v) < \frac{d(v)}{2}\right\} \end{aligned}$$

Ahora, notemos que  $d_{V_p(G)}(v) = \sum_{u \in N(v)} X_u =: X(v)$ , con  $X_u$  Bernoulli,  $\mathbb{P}\{X_u = 1\} = p$  y por ende  $\mathbb{E}(X(v)) = pd(v)$ . Luego,

$$\begin{aligned} \mathbb{P}\left\{d_{V_p(G)}(v) < \frac{d(v)}{2}\right\} &= \mathbb{P}\left\{X(v) < \frac{d(v)}{2}\right\} \\ &= \mathbb{P}\{X(v) < (1 - \delta)pd(v)\} \\ &= \mathbb{P}\{X(v) < (1 - \delta)\mathbb{E}(X(v))\}, \end{aligned}$$

aplicamos entonces la desigualdad de Chernoff (la versión presentada en [28]), obteniendo:

$$\begin{aligned} \mathbb{P}\left\{d_{V_p(G)}(v) < \frac{d(v)}{2}\right\} &< e^{-\frac{p\delta^2}{2}d(v)} \\ &\leq e^{-\frac{p\delta^2}{2}d}. \end{aligned}$$

Finalmente,

$$\begin{aligned} r_p(G) &> 1 - \sum_{v \in V_p(G)} e^{-\frac{p\delta^2}{2}d} \\ &\geq 1 - \sum_{v \in V} e^{-\frac{p\delta^2}{2}d} \\ &= 1 - ne^{-\frac{p\delta^2}{2}d} \end{aligned}$$

y se tiene el resultado. ■

Podemos entonces aplicar este lema a grafos en los cuales el grado mínimo es suficientemente grande. Diremos que una clase de grafos tiene grado mínimo  $d(n)$  si el grado mínimo de cualquier grafo con más de  $n$  vértices es al menos  $d(n)$ . Además, una secuencia de grafos  $(G_k)_{k \in \mathbb{N}}$  se dirá *creciente* si el número de vértices de los grafos, crece con  $k$ .

**Corolario 3.4.** Sean  $\frac{1}{2} < p < 1$  y  $d(n) = \omega(\log(n))$ . Entonces, para cada secuencia de grafos creciente  $(G_k)_{k \in \mathbb{N}}$  con grado mínimo  $d(n)$ , se tiene que  $\lim_{k \rightarrow \infty} r_p(G_k) = 1$ .

*Demostración.* Del cálculo anterior, si denotamos  $d_k$  al grado mínimo del grafo  $G_k$  y sin pérdida de generalidad suponemos que el número de vértices de  $G_k$  es  $k$  (ya que se trata de una secuencia creciente de grafos), tenemos:

$$\begin{aligned} r_p(G_k) &> 1 - ke^{-\frac{p\delta^2}{2}d_k} \\ &= 1 - e^{\log(k) - \frac{p\delta^2}{2}d_k} \\ &= 1 - e^{c \log(k) \left(1 - \frac{p\delta^2}{2} \frac{d_k}{c \log(k)}\right)}, \end{aligned}$$

con  $c > 0$ . Como la clase de grafos es de grado mínimo  $d(k)$  y  $d(k) = \omega(\log(k))$ , luego

$$r_p(G_k) \geq 1 - e^{c \log(k) \left(1 - \frac{p\delta^2}{2} \frac{d(k)}{c \log(k)}\right)} \rightarrow 1 \quad \text{cuando } k \rightarrow \infty. \quad \blacksquare$$

En el caso de que  $p > 0$  sea cualquiera, hay clases de grafos para las que se pueden obtener resultados similares. Tal es el caso de grafos regulares planares de grado máximo 4 y árboles.

## 3.2. Grafos planares regulares de grado 3 y 4

Sea  $G$  un grafo planar cúbico. Notemos que al considerar un ciclo  $C$  en el grafo y  $S = V(C)$ , como cada vértice en  $S$  tiene exactamente dos vecinos dentro del conjunto, teniendo por ende a lo más dos vecinos fuera de él, este conjunto resulta ser una comunidad de  $G$ .

Un tipo de ciclo particular es aquel de las fronteras de caras de  $G$ . El hecho de que estos sean suficientemente numerosos será de ayuda para estudiar  $r_p(G)$ .

Comenzamos estudiando el caso en que  $G$  es cúbico. Claramente, se tiene que  $3|V| = 2|E|$ . Denotemos por  $F$  al conjunto de las caras de  $G$ . Gracias a la fórmula de Euler, sabemos que  $|V| + |F| - |E| = 2$ . Luego, para grafos planares cúbicos se tendrá que  $|E| = 3|F| - 6$ . Sea ahora  $P_i$  el número de caras de tamaño  $i$  y  $P_{\leq k}$  el número de caras de tamaño a lo más  $k$ .

**Lema 3.5.** Sea  $G$  un grafo planar cúbico. Luego  $P_{\leq 6} \geq \frac{1}{12}|F|$ .

*Demostración.* Notemos primero que  $|F| = \sum_{i \geq 3} P_i$ , de donde se sigue que  $|E| = 3 \sum_{i \geq 3} P_i - 6$ . Por otra parte,  $\sum_{i \geq 3} iP_i = 2|E|$ . De estas dos ecuaciones, obtenemos:

$$6 \sum_{i \geq 3} P_i - 12 = \sum_{i \geq 3} iP_i \quad (3.1)$$

Simplificando, se tiene que:

$$3P_3 + 2P_4 + P_5 = 12 + \sum_{i \geq 7} (i - 6)P_i \quad (3.2)$$

Luego, revisamos dos posibles casos:

■ **Caso 1:**  $3P_3 + 2P_4 + P_5 \geq \frac{1}{4}|F|$ .

En este caso  $P_3 + P_4 + P_5 \geq \frac{1}{12}|F|$  y por ende  $P_{\leq 6} \geq \frac{1}{12}|F|$ .

■ **Caso 2:**  $3P_3 + 2P_4 + P_5 \leq \frac{1}{4}|F|$ .

En este caso, gracias a (3.2),  $\sum_{i \geq 7} P_i \leq \frac{1}{4}|F|$ .

Luego  $\sum_{i \neq 6} P_i \leq \frac{1}{2}|F|$  y así  $P_6 \geq \frac{1}{2}|F|$ . De donde se tiene que  $P_{\leq 6} \geq \frac{1}{12}|F|$ . ■

Con la ayuda de este lema, podemos concluir el siguiente resultado.

**Proposición 3.6.** *Dado  $p > 0$ , para cualquier secuencia creciente  $(G_k)_{k \in \mathbb{N}}$  de grafos cúbicos planares se tiene que  $\lim_{k \rightarrow \infty} r_p(G_k) = 1$ .*

*Demostración.* Sea  $G_k$  un grafo de la secuencia, con  $|V(G_k)| = n_k$ . Consideremos  $F_{disj}$  cualquier subconjunto maximal de caras disjuntas de tamaño menor o igual a 6. Como cada una de las caras en  $F_{disj}$  tiene a lo más 6 caras vecinas, se tiene que  $|F_{disj}| \geq |P_{\leq 6}| - 6|F_{disj}|$  y por ende, gracias al lema anterior:

$$|F_{disj}| \geq \frac{1}{7}|P_{\leq 6}| \geq \frac{1}{84}|F|$$

Ahora, tal como vimos antes, usando la fórmula de Euler y el hecho de que  $G_k$  es cúbico tenemos que  $|F| = 2 + \frac{|V|}{2}$ . Es decir, el número de caras del grafo crece linealmente con  $n_k$  y por ende  $|F_{disj}|$ .

Para cada cara  $f \in F_{disj}$ , los vértices de  $f$  son una comunidad de  $G_k$ . Luego

$$\begin{aligned} r_p(G_k) &\geq \mathbb{P}\{\exists f \in F_{disj} : V(f) \subseteq V_p(G_k)\} \\ &= \sum_{f \in F_{disj}} \mathbb{P}\{V(f) \subseteq V_p(G_k)\}, \end{aligned}$$

y la probabilidad de este último evento es al menos  $p^6$ , así

$$r_p(G_k) \geq |F_{disj}|p^6$$

lo cual vimos, crece linealmente con  $n_k$ . De aquí, se concluye el resultado. ■

Dado que los ciclos asociados a caras son también comunidades en grafos planares regulares de grado 4, el resultado se extiende de manera natural a dicha clase de grafos.

**Proposición 3.7.** *Dado  $p > 0$ , para cualquier secuencia creciente de grafos  $(G_k)_{k \in \mathbb{N}}$  de grafos planares regulares de grado 4, se tiene que  $\lim_{k \rightarrow \infty} r_p(G_k) = 1$ .*

Otra clase de grafos para la cual propiedades similares se manifiestan es aquella de los árboles.

### 3.3. Árboles

**Lema 3.8.** *Sea  $T_k$  un árbol con  $k$  vértices. Sea  $p > 0$  y  $q = 1 - p$ . Luego*

$$r_p(T_k) \geq (1 - q^{\frac{k}{2}})r_p(T_{k-1}).$$

*Demostración.* Sea  $k > 3$ . Sea  $V_2 \subseteq V(T_k)$  el conjunto de vértices en  $T_k$  con grado a lo más 2.

Es fácil ver que  $|V_2| \geq \frac{k}{2}$ . Probamos este hecho por inducción en la altura del grafo,  $h$  (medida desde una raíz  $r \in V(T_k)$ , fija). Para  $h = 1$ , sólo la raíz  $r$  tiene grado mayor a 2. Como el resto de los vértices son hojas y luego pertenecen a  $V_2$ , así  $|V_2| \geq k - 1 \geq \frac{k}{2}$ , ya que  $k > 3$ .

Ahora si  $T_k$  es de altura  $h > 1$ , consideramos  $U_h$  el conjunto de vértices de altura  $h$  y  $U_{<h}$ , aquel de los vértices de alturas menores que  $h$ . Se tiene que todos los vértices en  $U_h$  son hojas (porque el árbol es de altura  $h$ ) y por ende  $U_h \subseteq V_2$ . Si consideramos  $T := T_k \setminus U_h$ , que es un árbol de altura  $h - 1$ , se tendrá por hipótesis inductiva que  $|V_2(T)| \geq \frac{|U_{<h}|}{2}$  (con  $V_2(T)$  el conjunto de vértices de grado menor o igual a dos en  $T$ ). Cuando volvamos a agregar

los vértice de  $U_h$  al árbol  $T$ , es posible que vértices aumenten su grado de 1 a 3 o más, disminuyendo así el tamaño de  $V_2(T)$ . Sin embargo, estos vértices son a lo más uno por cada par de hojas de  $U_h$ . Resumiendo, obtenemos:

$$|V_2| \geq |V_2(T)| + |U_h| - \frac{|U_h|}{2} \geq \frac{|U_{<h}|}{2} + \frac{|U_h|}{2} = \frac{k}{2},$$

gracias a que  $|U_{<h}| + |U_h| = |U_{<h} \cup U_h| = |V(T_k)| = k$ . De aquí se concluye el resultado.

Veamos ahora que toda comunidad en  $T_k$  contiene al menos un vértice de  $V_2$ . En efecto, consideremos una comunidad  $S$ , de  $T_k$ . Sea  $v$  el vértice en  $S$  a mayor distancia en el árbol a la raíz  $r$ . Se tiene que  $v \in V_2$ , de lo contrario  $d_G(v) \geq 3$  y por ende debe tener al menos dos vecinos en  $S$  (para que la condición de comunidad se cumpla). Sin embargo uno de estos estará a una distancia mayor que  $v$  de  $r$ .

De esta manera, si  $V_p(T_k)$  no contiene vértice de  $V_2$ , este no podrá contener una comunidad de  $T_k$ . Se sigue entonces:

$$\begin{aligned} r_p(T_k) &\geq \mathbb{P}\{V_p(T_k) \text{ tiene una comunidad} \mid V_2 \cap V_p(T_k) \neq \emptyset\} \mathbb{P}\{V_2 \cap V_p(T_k) \neq \emptyset\} \\ &\quad + \mathbb{P}\{V_p(T_k) \text{ tiene una comunidad} \mid V_2 \cap V_p(T_k) = \emptyset\} \mathbb{P}\{V_2 \cap V_p(T_k) = \emptyset\} \\ &\geq \mathbb{P}\{V_p(T_k) \text{ tiene una comunidad} \mid V_2 \cap V_p(T_k) \neq \emptyset\} \mathbb{P}\{V_2 \cap V_p(T_k) \neq \emptyset\} + 0 \\ &\geq r_p(T_{k-1})(1 - q^{|V_2|}), \end{aligned}$$

y ya que  $|V_2| \geq \frac{k}{2}$  y  $p < 1$ ,

$$\geq r_p(T_{k-1})(1 - q^{\frac{k}{2}}) \quad \blacksquare$$

**Proposición 3.9.** *Dado  $p > 0$ , existe  $\delta_p > 0$  tal que para todo árbol  $T$  se tiene que  $r_p(T) \geq \delta_p$ .*

*Demostración.* Notemos primero que  $r_p(T_3) = (1 - q)(1 - q^2)$ , es decir, la probabilidad de que el vértice de grado 2 esté en  $V_p(T)$  y que alguno de los vértices de grado 1 también lo haga. Luego, tenemos que  $r_p(T_3) \geq (1 - q^{\frac{1}{2}})(1 - q^{\frac{2}{2}})(1 - q^{\frac{3}{2}})$ . Así, por el lema previo:

$$r_p(T_k) \geq \prod_{i=1}^k (1 - q^{\frac{i}{2}}).$$

Luego

$$\begin{aligned}
 r_p(T_k) &\geq \exp\left(\sum_{i=1}^k (\ln(1 - q^{\frac{i}{2}}))\right) \\
 &\geq \exp\left(\sum_{i=1}^k 1 - \frac{1}{1 - q^{\frac{i}{2}}}\right) \\
 &= \exp\left(\sum_{i=1}^k \frac{-q^{\frac{i}{2}}}{1 - q^{\frac{i}{2}}}\right) \\
 &\geq \exp\left(-\frac{1}{1 - q} \sum_{i=1}^k q^{\frac{i}{2}}\right) \\
 &\geq \exp\left(-\frac{1}{(1 - q)(1 - \sqrt{q})}\right) = \delta_p > 0
 \end{aligned}$$

■

### 3.4. Conjeturas

En este contexto, se conjetura que si  $p$  es grande es posible encontrar comunidades en  $V_p(G)$  con una probabilidad positiva dada, para todo grafo. Es decir:

**Conjetura 1.** *Sea  $p > \frac{1}{2}$ . Luego existe  $\delta_p > 0$  tal que para todo grafo  $G$ , se tiene que  $r_p(G) \geq \delta_p$ .*

Además, si el grado máximo del grafo está controlado, se conjetura un resultado aun más fuerte:

**Conjetura 2.** *Sea  $d \in \mathbb{N}$  y  $p > 0$ . Dada cualquier secuencia creciente de grafos  $(G_k)_{k \in \mathbb{N}}$  de grado máximo a lo más  $d$ , se tiene que  $\lim_{k \rightarrow \infty} r_p(G_k) = 1$ .*

# Capítulo 4

## El Coeficiente de Comunidad

### 4.1. Un primer acercamiento

Dado un vértice  $v$  de un grafo  $G = (V, E)$ , sea  $S_v \subseteq V$  una comunidad de tamaño mínimo conteniendo a  $v$ . Una observación que motiva esta parte del trabajo es: en grafos localmente estructurados el valor de  $|S_v|$  es usualmente pequeño.

De esta manera, a modo de medir qué tan localmente estructurado es un grafo alrededor de  $v$ , surge el deseo de asociar a cada  $v$  el valor  $\frac{|S_v|}{n}$ . Sin embargo, como veremos a continuación, calcular el tamaño de  $S_v$  resulta ser NP-duro. Formalicemos esto, definiendo el problema COMUNIDAD.

COMUNIDAD

**Instancia:** Grafo  $G$  y  $k \in \mathbb{N}$

**Pregunta:** ¿Hay alguna comunidad  $S$  en  $G$  tal que  $|S| \leq k$ ?

Para probar la NP-completitud de COMUNIDAD, usaremos el hecho de que el siguiente problema es NP-completo [2].

HALF-CLIQUE

**Instancia:** Grafo  $G$ , con  $n$  vértices.

**Output:** ¿Hay algún clique en  $G$  de tamaño al menos  $\frac{n}{2}$ ?

**Proposición 4.1.** HALF-CLIQUE es NP-completo.

*Demostración.* Se tiene que CLIQUE se reduce a HALF-CLIQUE. La reducción es la siguiente, dada una instancia de CLIQUE,  $(G, k)$  en donde  $G$  tiene  $n$  vértices; sean los conjuntos  $V_n$  y  $V_{2k}$  de  $n$  y  $2k$  nuevos vértices respectivamente. Se agregan estos dos conjuntos a los vértices de  $G$  y se conecta cada vértice de  $V_n$  a todo vértice de  $G$ . Además, se conectan todos los vértices dentro de  $V_n$ . Llamamos a este nuevo grafo  $G'$ .

Se tiene claramente que  $G$  tiene un clique de tamaño  $k$  si y sólo si  $G'$ , tiene uno de tamaño  $k + n$  (los cliques de tamaño  $k + n$  de  $G'$  son  $V_n$  y un clique de tamaño  $k$  de  $G$ ). Pero  $G'$  tiene exactamente  $2(k + n)$  vértices, por lo que lo último es una instancia de HALF-CLIQUE. ■

Ahora podemos probar que COMUNIDAD es NP-completo, reduciendo HALF-CLIQUE a COMUNIDAD.

**Proposición 4.2.** COMUNIDAD es NP-completo.

*Demostración.* Sea  $G = (V, E)$  una instancia de HALF-CLIQUE. Construimos el grafo  $G'$  de la siguiente manera (ver Figura 4.1). Consideramos los grafos  $G_1 = (V, E)$ ,  $G_2 = (V, E)$  y  $V_1, V_2$  dos copias de  $V$ . Además, consideramos cuatro vértices nuevos,  $u_1, u_2, v_1, v_2$ . Ahora, para  $i \in \{1, 2\}$ , se conecta cada vértice  $v$  de  $G_i$  con aquellos vértices  $u \in V_i$  tales que  $v \in V \setminus (\{v\} \cup N_G(u))$ . Además, el vértice  $u_i$  se conecta con todos los vértices en  $G_i$  y  $v_i$  con todos los de  $V_i$ . Finalmente, se conectan los vértices de  $V_1$  con aquellos de  $V_2$  según en  $G$ , es decir  $v \in V_1$  se conecta con los vértices  $u \in V_2$  tales que  $v \in N_G(u)$ .

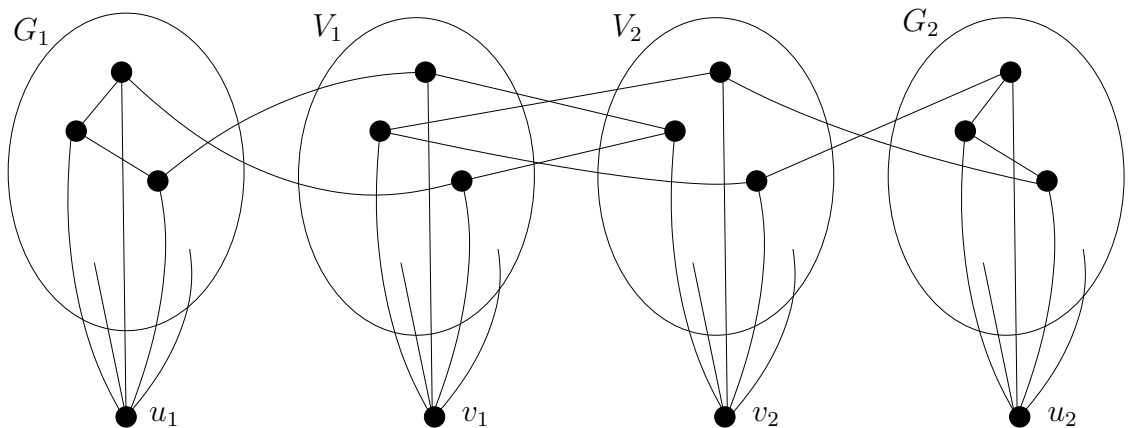


Figura 4.1: Esquema de la reducción.

Claramente se tiene que  $G'$  es un grafo  $n$ -regular.

Ahora, cualquier clique  $C$  de  $G$ , de tamaño al menos  $\frac{n}{2}$ , induce un clique de  $G'$  de tamaño  $\lceil \frac{n}{2} \rceil + 1$ , para esto basta considerar  $C' = C \cup \{u_1\}$ . Este conjunto resulta ser además una

comunidad en  $G'$ , ya que para un vértice  $v \in C'$  se tiene que:

$$d_{C'}(v) = \left\lceil \frac{n}{2} \right\rceil = \left\lceil \frac{d_{G'}(v)}{2} \right\rceil.$$

Por otra parte, si  $S \subseteq V(G')$  es una comunidad de  $G'$  de tamaño  $\left\lceil \frac{n}{2} \right\rceil + 1$ , es también un clique de  $G'$  y se tendrá que

$$S \subseteq V(G_1) \cup V_1 \cup \{u_1, v_1\} \quad \text{ó} \quad S \subseteq V(G_2) \cup V_2 \cup \{u_2, v_2\},$$

ya que no hay adyacencia entre vértices de  $G_1 \cup \{u_1\}$  y  $G_2 \cup \{u_2\}$ ,  $V_1 \cup \{v_1\}$  y  $G_2 \cup \{u_2\}$  o  $V_2 \cup \{v_2\}$  y  $G_1 \cup \{u_1\}$ . Además porque como  $S$  puede tener a lo más un vértice de  $V_1$  o  $V_2$  (ya que ambos son por sí solos conjuntos independientes), si  $S \subseteq V_1 \cup V_2$ , entonces  $S$  está formado por los extremos de una arista. Pero suponemos que  $n > 2$ , por lo que esto no ocurre (para el caso  $n = 2$ , el clique inducido en  $G$  es simplemente un vértice).

Por la misma razón (independencia de  $V_1$  y  $V_2$ ),  $S$  no contiene a  $v_1$  ni  $v_2$ . Entonces s.p.g.,  $S$  está conformado por  $\left\lceil \frac{n}{2} \right\rceil$  vértices en  $G_1$  y a lo más un vértice en  $V_1$  o a  $u_1$ , no pudiendo contener ambos ya que  $u_1$  no es adyacente a ningún vértice en  $V_1$ . Así,  $S \cap V(G_1)$  es un clique de tamaño al menos  $|S| - 1 = \left\lceil \frac{n}{2} \right\rceil$ .

De esta manera,  $G \in \text{HALF-CLIQUE}$  ssi  $(G', \left\lceil \frac{n}{2} \right\rceil + 1) \in \text{COMUNIDAD}$ . Obteniéndose así el resultado. ■

## 4.2. Definiendo $cc(G)$

Pese a que el resultado previo implica que en la práctica no hay una forma eficiente de calcular  $|S_v|$ , es posible realizar una modificación a la medida propuesta. Si buscamos una medida de “clustering”, bastará calcular la *profundidad* de  $v$ , definida como la mínima distancia que es necesario alejarse de  $v$  para encontrar una comunidad conteniendo  $v$ :

**Definición.** Sea un grafo  $G = (V, E)$  y  $v \in V$ , se define

$$depth_G(v) = \text{mín}\{ecc_S(v) : S \text{ es comunidad con } v \in S\}$$

en donde  $ecc_S(v)$  es la excentricidad de  $v$  en  $S$ , la distancia desde  $v$  al vértice más lejano en  $S$ .

Además, normalizando esta cantidad, definimos el *coeficiente de comunidad* como:

$$cc_G(v) = \frac{depth_G(v)}{ecc_G(v)}.$$

Finalmente, el coeficiente de comunidad del grafo  $G$  está definido como el promedio de  $cc_G(v)$  tomado sobre todos los vértices. Es decir,  $cc(G) = \frac{1}{n} \sum_{v \in V} cc_G(v)$ .

A continuación veremos que  $cc(G)$  es de hecho calculable en tiempo polinomial.

### 4.3. Calculando $cc(G)$

Para calcular la profundidad de un vértice necesitamos definir la *dinámica de regla de mayoría fuerte e irreversible* en un grafo  $G = (V, E)$  [22]. Sea  $S \subseteq V$  y a partir de él definimos una configuración inicial como la asignación de 0s y 1s a los vértices de  $V$  de la siguiente manera: asignamos un 0 a  $v$  si  $v \in S$  y 1 de otra manera. La dinámica es la siguiente: si un vértice tiene estrictamente más 1s que 0s en su vecindad, luego se convierte en 1 (y 1s nunca se transforman en 0s). El sistema llega al mismo estado de punto fijo  $FP(S)$ , si la regla se aplica tanto en serie como en paralelo. Denotemos  $m(S) \subseteq V$  al conjunto de vértices que se encuentran en estado 0 en  $FP(S)$ . En otras palabras,  $m(S)$  corresponde a aquellos vértices que “no han sido contagiados por los 1s”.

Veamos primero algunas propiedades de la dinámica:

**Proposición 4.3.** *Sea  $\phi \neq S' \subseteq S \subseteq V$ . Se tiene que*

1.  $m(S') \subseteq m(S)$ .
2.  $S'$  es una comunidad  $\Rightarrow S' \subseteq m(S)$ .
3.  $S$  es una comunidad  $\Leftrightarrow m(S) = S$ .
4.  $m(S) \neq \phi \Leftrightarrow m(S)$  es una comunidad.
5.  $m(S) = \phi \Leftrightarrow S$  no contiene comunidades.

*Demostración.* Notemos que en la primera aplicación de la regla de mayoría, los vecinos de cada vértice en  $S$  son al menos los vecinos que posee en  $S'$ . Luego si un vértice de  $S'$  no cambia a estado 1, tampoco lo hará en  $S$ . Este mismo argumento se tiene para aplicaciones sucesivas de la regla de mayoría. Por ende,  $m(S') \subseteq m(S)$ . Esto prueba 1.

Ahora, si  $S'$  es comunidad todos sus vértices tienen al menos la mitad de vecinos en  $S'$ , por lo que no cambiarán nunca de estado y luego  $S' \subseteq m(S')$ . Y como  $m(S') \subseteq m(S)$ , se tiene que  $S' \subseteq m(S)$ . Se prueba así 2 y además que  $S' \subseteq m(S')$ , si  $S'$  es comunidad.

Probamos ahora 3. Se tiene que todo vértice en  $m(S)$  necesariamente pertenece a  $S$  en la condición inicial, ya que no es posible que vértices de estado 0 cambien a estado 1 y regresen a estado 0 durante la dinámica. Así  $m(S) \subseteq S$ . Gracias a esto y a la observación hecha en el párrafo anterior,  $m(S) = S$  si  $S$  es comunidad.

Por otra parte, si se tiene que  $m(S) = S$ , al encontrarnos en el punto fijo de la dinámica una nueva aplicación de la regla de mayoría no debe afectar el estado de ningún vértice en  $m(S)$ . Para que esta condición se satisfaga, cada vértice en  $m(S)$  debe tener al menos la misma cantidad de vecinos dentro que fuera del conjunto. Esto significa precisamente que  $m(S) = S$  es una comunidad.

Por la misma razón anterior, cada vez que  $m(S) \neq \phi$ ,  $m(S)$  es una comunidad. Probándose así 4.

Finalmente gracias a 2, toda comunidad contenida en  $S$  está también contenida en  $m(S)$ . Por esto si  $m(S) = \phi$ ,  $S$  no contiene comunidades. ■

Esta última proposición permite diseñar un algoritmo que calcula  $depth_G(v)$ . Se comienza con  $S = \{v\} \cup N_G(v)$ . Si  $m(S) \neq \phi$  luego se tiene que  $m(S)$  es una comunidad y por ende  $depth_G(v) = 1$ . De lo contrario,  $depth_G(v) > 1$ , y redefinimos  $S = S \cup N_G(S)$ .

El algoritmo DEPTH se define formalmente como:

```

DEPTH Input:  $G = (V, E), v \in V$ . Output:  $depth(v)$ .
 $S \leftarrow \{v\}, r \leftarrow 0$ 
while  $r \leq n$  do
  Encontrar  $m(S)$  (aplicando la regla de mayoría)
  if  $m(S) \neq \phi$  then
    return  $r$ 
  end if
   $S \leftarrow S \cup N_G(S)$ 
   $r \leftarrow r + 1$ 
end while

```

Se tiene que

**Proposición 4.4.** *DEPTH* retorna  $depth_G(v)$  y su complejidad en tiempo es  $O(n^3)$ .

*Demostración.* Veamos primero la correctitud del algoritmo. Notemos primero que  $V$  en sí mismo es una comunidad. Sea  $d^* = depth_G(v)$  y sea  $S^*$  una comunidad conteniendo a  $V$  y tal que  $ecc_{S^*}(v) = d^*$ . En cada paso  $k$  el conjunto  $S$  es el conjunto

$$N_G^k(v) = \{u \in V : d_G(v, u) \leq k\}.$$

Si  $k < d^*$ , se tendrá que  $m(N_G^k(v)) = \phi$ . De lo contrario  $depth_G(v)$  sería más pequeño. Por otra parte, como  $S^* \subseteq N_G^{d^*}(v)$ , por la proposición anterior se tiene que  $m(N_G^{d^*}(v)) \neq \phi$ . Es decir el algoritmo retornará  $d^*$ .

Para la complejidad, notamos que calcular  $m(S)$ , para cada valor de  $r$ , toma a lo más  $n^2$  pasos. Esto porque la regla de mayoría se aplica a lo más  $n$  veces (al menos un 0 se convierte en 1, cada vez) en a lo más  $n$  vértices. Como  $r$  puede tomar valores entre 0 y  $n$ , se obtiene el tiempo  $O(n^3)$ . ■

Ahora, para completar el cálculo del coeficiente, se debe calcular la excentricidad de todo vértice  $v \in V(G)$ . Esto puede hacerse corriendo  $n$  veces el algoritmo de *Dijkstra* en cada vértice del grafo (ya que la distancia de los arcos es positiva), el cual con una implementación ingenua, toma  $O(n^2)$ . Esta complejidad puede ser mejorada con implementaciones usando *heaps*, para ello ver [29].

Así, finalmente se tiene que

**Proposición 4.5.** *El problema de calcular  $cc(G)$  tiene complejidad de tiempo  $O(n^4)$ .*

*Demostración.* Para cada vértice  $v \in V(G)$ , se corre DEPTH para calcular  $depth_G(v)$  y luego alguna implementación de Dijkstra. Todo esto toma  $O(n^4)$  (lo cual incluye promediar sobre todo vértice, que toma  $O(n)$ ). ■

# Capítulo 5

## Resultados Experimentales

Dado que el coeficiente de comunidad puede ser calculado eficientemente, en este capítulo realizamos experimentos para estudiar su comportamiento en dos clases de grafos,  $\alpha$ -Grafos y  $\beta$ -Grafos. Estos grafos han sido usados anteriormente para estudiar al coeficiente de clustering [7, 6]. Gracias a esto, compararemos ambos coeficientes.

Las simulaciones se realizaron en Python, usando el paquete NetworkX [30] para manipulación de redes complejas.

### 5.1. $\alpha$ -Grafos

El modelo  $\alpha$  se construye agregando aristas a un substrato dado (que es en este caso un anillo, debido a las características de ser una estructura minimal conexa y de diámetro grande) según la cantidad de vecinos en común de vértices. En concreto, la probabilidad de agregar una arista  $(i, j)$  al grafo es proporcional a la cantidad  $R_{ij}$ , dada por:

$$R_{ij} = \begin{cases} 1, & \text{Si } m_{ij} \geq k \\ \left[\frac{m_{ij}}{k}\right]^\alpha (1 - p) + p, & \text{Si } 0 < m_{ij} < k \\ p, & \text{Si } m_{ij} = 0 \end{cases}$$

En donde  $m_{ij}$  es el número de vecinos en común entre  $i$  y  $j$ . Además,  $p$  es una probabilidad base de conexión entre vértices y  $k$  es el grado esperado. Este modelo favorece la conexión entre vértices con gran cantidad de vecinos en común.

Se presenta en la Figura 5.1 y la Figura 5.2 los resultados de calcular  $cc(G)$  y  $c(G)$  para diversos valores de  $\alpha$ , promediados sobre 20 realizaciones del grafo. Para la probabilidad base, se usó  $p = 10^{-10}$ .

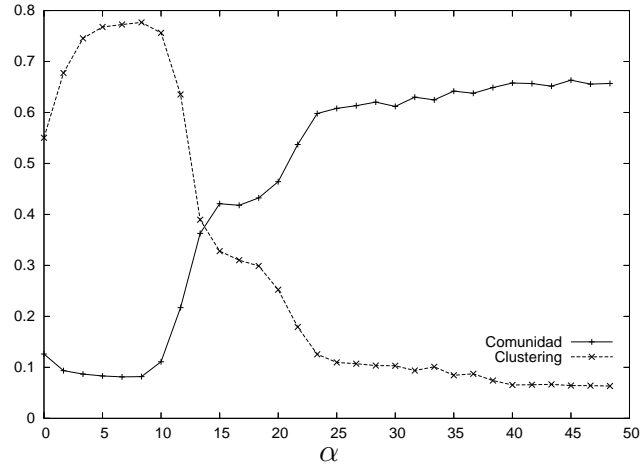


Figura 5.1:  $cc(G)$  vs.  $c(G)$ . Modelo  $\alpha$ , con  $n = 100$  y  $k = 6$ .

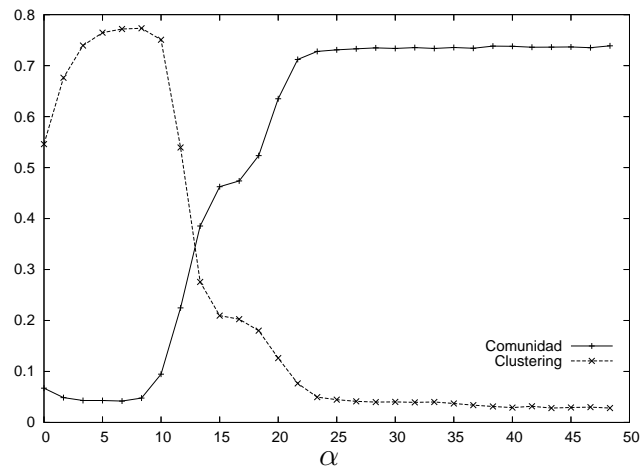


Figura 5.2:  $cc(G)$  vs.  $c(G)$ . Modelo  $\alpha$ , con  $n = 200$  y  $k = 6$ .

Notamos que el comportamiento del coeficiente de comunidad es similar al del coeficiente de clustering, en cuanto a discriminar grafos locales (generados con  $\alpha$  pequeño) y grafos cerca del límite aleatorio (para  $\alpha$  grande). Naturalmente dicha discriminación se produce de manera opuesta para el coeficiente de comunidad, debido a que mientras menos local sea el grafo, más grande es el radio de una comunidad minimal en torno a un vértice.

Otra observación es que el “salto” en el coeficiente de comunidad (la transición desde valores  $\alpha$  pequeño hacia valores de  $\alpha$  grandes), se produce para los mismos valores del parámetro  $\alpha$  que en el caso del coeficiente de clustering. Lo cual permite usar el coeficiente de

comunidad para reconocer el fenómeno “pequeño mundo”, es decir coeficiente de clustering alto y diámetro pequeño.

Además, se tiene que para valores de  $\alpha$  grandes, es decir cerca del límite aleatorio,  $cc(G)$  se estabiliza tal como lo hace  $c(G)$  (el cual debería cumplir que  $\lim_{\alpha \rightarrow \infty} c(G) \sim (k/n)$ , como se muestra en [6]).

## 5.2. $\beta$ -Grafos

El modelo  $\beta$  consiste en un 1-lattice perfecto, en donde cada vértice tiene  $k$  vecinos ( $k/2$  a cada lado), y luego se *reconectan aleatoriamente* las aristas, con probabilidad  $\beta$ . Este procedimiento se realiza escogiendo un vértice y una arista que lo conecte con su vecino más cercano en sentido horario. Con probabilidad  $\beta$ , se reconecta este extremo a otro vértice escogido aleatoriamente. Este procedimiento se repite para todos los vértices del anillo. Luego se realiza otra vuelta por el anillo, reconectando aristas hacia segundos vecinos más cerca en el sentido horario. Así continuamos el proceso, considerando vecinos cada vez más lejanos, hasta que toda arista es considerada una vez para reconexión. Un esquema de la forma de estos grafos para diversos valores de  $\beta$  se presenta en la Figura 5.3.

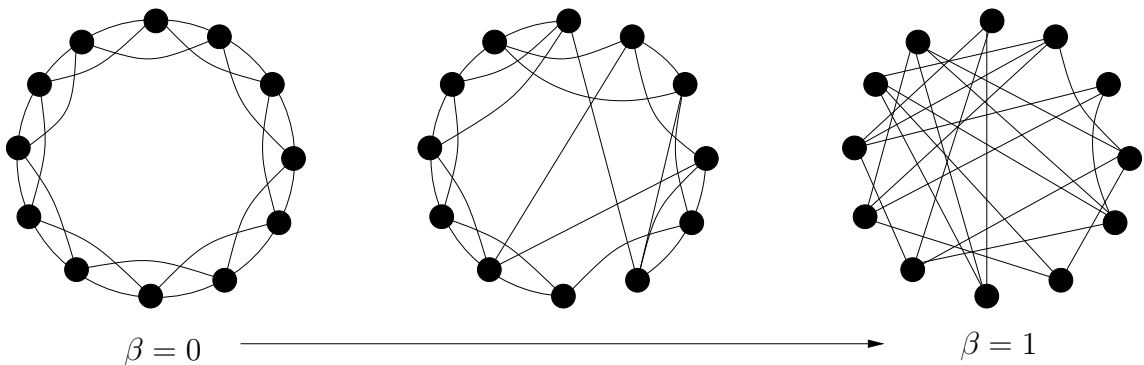


Figura 5.3:  $\beta$ -Grafo, para diversos valores de  $\beta$ .

En la Figura 5.4 y la Figura 5.5 se observa el valor de ambos coeficientes para  $0 \leq \beta \leq 1$ . Podemos observar que el comportamiento de ambos coeficientes es similar, en el sentido de discriminar los extremos de grafos con clustering alto ( $\beta$  pequeño) y cerca del extremo aleatorio ( $\beta \sim 1$ ). De la misma manera que en el modelo  $\alpha$ ,  $cc(G)$  se estabiliza en el extremo aleatorio. En resumen, observamos que los coeficientes de comunidad y clustering tienen comportamientos similares. Además, un coeficiente de comunidad bajo está asociado a grafos localmente estructurados.

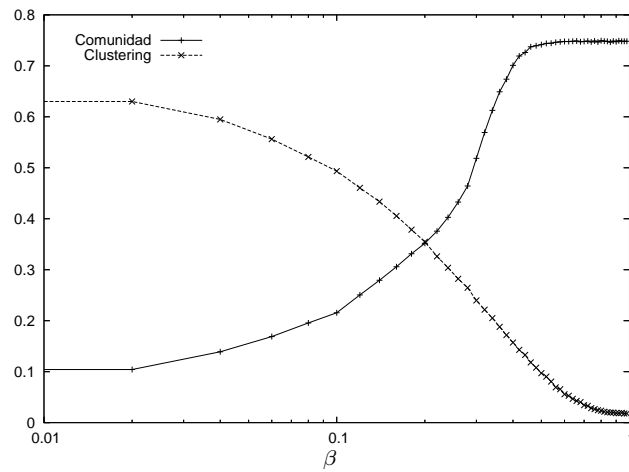


Figura 5.4:  $cc(G)$  vs.  $c(G)$ . Modelo  $\beta$ , con  $n = 500$  y  $k = 6$ . Se utiliza escala semilogarítmica.

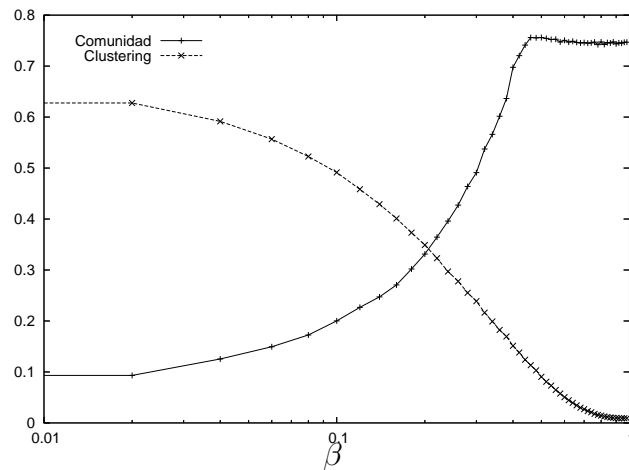


Figura 5.5:  $cc(G)$  vs.  $c(G)$ . Modelo  $\beta$ , con  $n = 1000$  y  $k = 6$ . Se utiliza escala semilogarítmica.

## Capítulo 6

# El Coeficiente de Comunidad en Grafos Planares sin Puentes Regulares de Grado 3 y 4

Una primera observación evidente es que en la grilla de  $m \times n$  vértices, el coeficiente de comunidad se va a cero, al crecer el tamaño del grafo. De hecho en este caso todo vértice pertenece a una comunidad pequeña: un ciclo de largo 4. El objetivo de este capítulo es intentar generalizar en lo posible este resultado.

Consideramos  $G = (V, E)$  un grafo planar cúbico. Si pensamos en  $G$  ya dibujado en el plano (al ser planar), luego los ciclos correspondientes a fronteras de caras son comunidades y de hecho entre ellos se encuentran las comunidades minimales de vértices. Llamaremos a tales comunidades *F-comunidades* y denotaremos al conjunto de F-comunidades por  $F(G)$ . Una función  $f : V \rightarrow F(G)$ , será llamada *F-función*. Una F-función particular que llamamos  $f_{min}$  es aquella que asocia a cada vértice la cara de largo menor que lo contiene, es decir

$$f_{min}(v) = \arg \min_{f \ni v} |f|.$$

De esta manera, la excentricidad de la comunidad más pequeña en torno a  $v$  será el diámetro de la cara más pequeña entre las que lo contienen, es decir

$$depth_G(v) = \left\lceil \frac{|f_{min}(v)|}{2} \right\rceil$$

Calculamos primero el coeficiente para un grafo con radio pequeño entre los cúbicos planares. Consideramos entonces el grafo formado por tres árboles binarios completos de profundidad  $k - 1$  (que llamaremos subárboles), unidos a un vértice central, y por un ciclo conectando a todas las hojas de los subárboles (ver Figura 6.1).

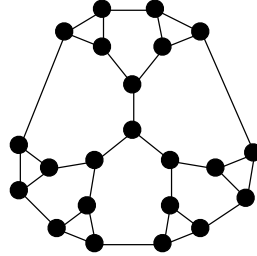


Figura 6.1: Grafo cúbico formado a partir de árboles binarios completos de profundidad  $k - 1 = 2$ .

Notemos que el número de vértices de este grafo es  $n = 3 \left( \sum_{i=0}^{k-1} 2^i \right) + 1 = 3 \cdot 2^k - 2$ , de donde

$$k = \log\left(\frac{n+2}{3}\right).$$

Además se tiene que el radio del grafo, es decir la mínima excentricidad, es igual a  $k$  (la excentricidad del vértice raíz).

Usamos la misma notación que en la Sección 3.2. Es posible deducir además que para grafos planares cúbicos se tiene que  $|E| = 3|F| - 6$  y  $|F| = n/2 + 2$ . Notando que  $|F| = \sum_{i \geq 3} P_i$  y además  $2|E| = \sum_{i \geq 3} iP_i$ , de lo anterior se obtiene

$$\sum_{i \geq 3} iP_i = 6 \sum_{i \geq 3} P_i - 12 \tag{6.1}$$

Ahora, calculamos el coeficiente de comunidad:

$$cc(G) = \frac{1}{n} \sum_{v \in V} \frac{depth_G(v)}{ecc(v)}.$$

Como la excentricidad mínimas es  $k$  y  $depth_G(v) \leq |f_{min}(v)|$ ,

$$cc(G) \leq \frac{1}{nk} \sum_{v \in V} |f_{min}(v)|.$$

Calcularemos explícitamente el valor de  $\sum_{v \in V} |f_{min}(v)|$ , notando que el vértice central del grafo pertenece a las tres caras de largo  $2k + 1$  y los  $3 \cdot 2^{k-1}$  vértices en la frontera del grafo (hojas de los subárboles) pertenecen a caras de largo 3 y por ende su valor de  $|f_{min}(v)|$  es 3. Por

otra parte, los vértices interiores de los subárboles tienen como cara de menor largo a aquella cara a la que pertenecen sus dos vecinos en un nivel más de profundidad en el subárbol (ver Figura 6.2), cuyo largo es  $2(k - i) + 1$  para un vértice de profundidad  $i$  (en donde suponemos que las raíces de los subárboles tienen profundidad 1) y se tiene que hay  $2^{i-1}$  vértices de dicha profundidad. Multiplicando por los tres subárboles las cantidades correspondientes, se obtiene:

$$\sum_{v \in V} |f_{\min}(v)| = (2k + 1) + 9 \cdot 2^{k-1} + 3 \left( \sum_{i=1}^{k-1} 2^{i-1} (2(k - i) + 1) \right)$$

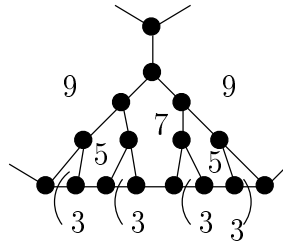


Figura 6.2: Caras y su largo en un subárbol. Caso  $k = 4$ .

Por otra parte, calculamos  $\sum_{i \geq 3} iP_i$ . Tenemos la cara exterior del grafo, de largo  $3 \cdot 2^{k-1}$  (la única de dicho largo); además hay tres caras de largo  $2k + 1$ , ubicadas entre los subárboles del grafo.

Finalmente, debemos contar las caras interiores restantes. En cada subárbol, por cada uno de los  $2^{i-1}$  vértices interiores de profundidad  $i$ , hay una cara de largo  $2(k - i) + 1$  asociada (aquella a la que pertenecen sus hijos en el subárbol). Luego, el conteo resulta:

$$\sum_{i \geq 3} iP_i = 3(2k + 1) + 3 \cdot 2^{k-1} + 3 \left( \sum_{j=1}^{k-1} 2^{j-1} (2(k - i) + 1) \right)$$

De esta manera, se concluye que

$$\begin{aligned} \sum_{v \in V} |f_{\min}(v)| &= \sum_{i \geq 3} iP_i + 3 \cdot 2^k - (4k + 2) \\ &= \sum_{i \geq 3} iP_i + 3 \cdot 2^{\log(\frac{n+2}{3})} - (4k + 2) \\ &= \sum_{i \geq 3} iP_i + n - 4k. \end{aligned}$$

Volviendo, entonces al cálculo de  $cc(G)$ , tenemos

$$cc(G) = \frac{1}{nk} \left( \sum_{i \geq 3} iP_i + n - 4k \right)$$

y utilizando la Ecuación (3.1),

$$\begin{aligned} cc(G) &= \frac{1}{nk} \left( 6 \sum_{i \geq 3} P_i - 12 + n - 4k \right) \\ &= \frac{1}{nk} (6|F| - 12 + n - 4k). \end{aligned}$$

Además como  $n = 2|F| - 4$  y en cualquier grafo cúbico planar  $|F| \geq 4$ , se tiene que  $n \geq |F|$  y por ende,

$$\begin{aligned} cc(G) &\leq \frac{6|F|}{|F|k} - \frac{12}{nk} + \frac{1}{k} - \frac{4}{n} \\ &= \frac{7}{k} - \frac{12}{nk} - \frac{4}{n} \\ &= \frac{7}{\log((n+2)/3)} - \frac{12}{n \log((n+2)/3)} - \frac{4}{n} \end{aligned}$$

De aquí se deduce que  $cc(G) \xrightarrow{n \rightarrow \infty} 0$ .

El cálculo anterior motiva la sospecha que todo grafo cúbico planar que tenga un radio que crezca con  $n$ , tendrá un coeficiente de comunidad que se va a 0 con  $n$ . La razón por la que esto sucedería es que el tamaño promedio de un comunidad en torno a un vértice (un ciclo o cara) es  $O(1)$ .

Confirmaremos parcialmente esta sospecha, probando que todo grafo planar cúbico pero sin puentes<sup>1</sup>, cumple que  $\lim_{n \rightarrow \infty} cc(G) = 0$ . Primero veremos que el radio de un grafo de grado máximo dado crece de manera al menos logarítmica en  $n$ . Luego, como  $diam(G) \leq 2 \cdot r(G)$ , el radio también crecerá al menos de manera logarítmica con  $n$ .

El siguiente resultado, relacionado con el problema introducido por E.F. Moore de acotar el número de vértices de un grafo de diámetro  $D$  y grado máximo  $\Delta$  dados [31] ( $n_{\Delta, D}$ ), permite deducir el crecimiento buscado.

**Teorema 6.1.** [E.F Moore, 1958]

$$n_{\Delta, D} \leq 1 + \Delta + \Delta(\Delta - 1) + \dots + \Delta(\Delta - 1)^{D-1}$$

---

<sup>1</sup>Un puente es una arista cuya remoción desconecta a  $G$ .

y por ende,

$$\begin{aligned} n_{\Delta,D} &\leq \frac{\Delta(\Delta-1)^D - 2}{\Delta-2} \quad \text{para } \Delta \geq 3 \\ n_{2,D} &\leq 2D + 1 \end{aligned}$$

*Demostración.* Ver [32]. ■

Así, se tiene el siguiente corolario:

**Corolario 6.2.** *Si  $G$  es cúbico,  $r(G) = \Omega(\log(n))$*

*Demostración.* Gracias al Teorema 6.1,

$$n \leq 3 \cdot 2^{\text{diam}(G)} - 2 \Leftrightarrow \log\left(\frac{n+2}{3}\right) \leq \text{diam}(G)$$

además, como  $\text{diam}(G) \leq 2 \cdot r(G)$ , se tiene el resultado. ■

A continuación veremos un par de lemas relacionados con la existencia de una F-función (que puede no ser  $f_{\min}$ ), de suma importancia para probar el resultado.

**Lema 6.3** ([33]). *Sea  $G = (V, E)$  un grafo plano cúbico y sin puentes. Luego existe una F-función  $f_0$  tal que para toda F-comunidad  $h \in F(G)$ ,  $|f_0^{-1}(h)| \leq 5$ .*

*Demostración.* Consideremos el grafo dual de  $G$ ,  $G^* = (F(G), E^*)$ . Como  $G$  es plano, cúbico y no tiene puentes, luego  $G^*$  es un grafo plano sin bucles ni aristas múltiples. Buscamos una asignación  $f_0$ , que asocie a cada vértice una F-comunidad particular a la que pertenezca. Por dualidad (ver Figura 6.3), esto es equivalente a buscar una función  $f^*$  que asigna a cada cara  $h^* \in G^*$  un vértice particular  $v^* \in G^*$  con  $v^*$  en la frontera de  $h^*$ . Luego, debemos probar que en la construcción de  $f^*$  a lo más 5 caras de  $G^*$  reciben el mismo vértice en la asignación.

Procedemos por inducción en el número de vértices de  $G^*$ . Si  $G^*$  tiene un vértice, la única cara del grafo es marcada con dicho vértice. Supongamos ahora que  $G^*$  tiene  $n + 1$  vértices. Un hecho clave que usaremos a continuación es

*Todo grafo planar simple (sin bucles ni aristas múltiples) tiene al menos un vértice de grado a lo más 5.*

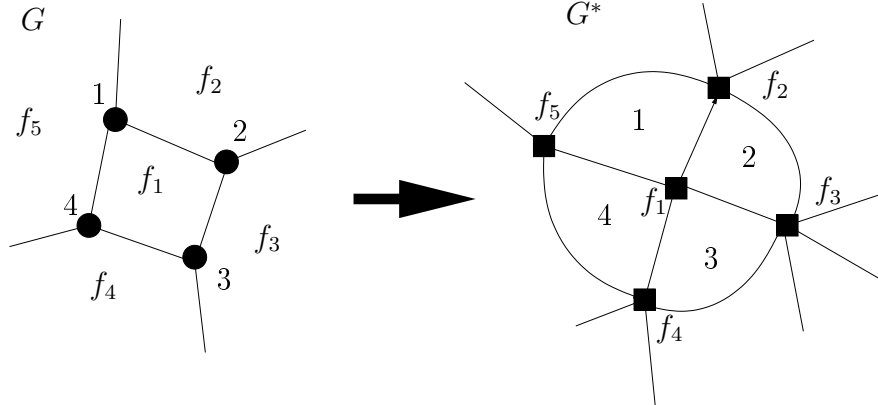


Figura 6.3: Ejemplo de parte del grafo y su dual.

Si esto no fuese cierto, tendríamos que  $|E^*| \geq \frac{6}{2}|V^*|$ , contradiciendo la relación entre aristas y vértices que se satisface para este tipo de grafos:  $|E^*| \leq 3|V^*| - 6$ .

Sea entonces  $v^*$  un vértice tal. Consideramos el grafo  $G' := G^* \setminus v^*$  (i.e. borramos el vértice y todas las aristas incidentes). Por hipótesis de inducción podemos resolver el problema de asignación en  $G'$ . El lugar ocupado por  $v^*$  corresponde a una cara de  $G'$  y dicha cara contiene a lo más 5 caras en  $G^*$ . Asignamos  $v^*$  a todas estas caras, obteniendo el resultado. ■

De aquí es posible acotar la suma de las comunidades en torno a los vértices de  $G$ .

**Lema 6.4.** *Sea  $G = (V, E)$  un grafo planar cúbico sin puentes. Luego existe una F-función  $f_0$  tal que  $\sum_{v \in V} |f_0(v)| \leq 15|V|$ .*

*Demostración.* Consideremos la F-función  $f_0$  del lema previo. Luego:

$$\sum_{v \in V} |f_0(v)| = \sum_{h \in F(G)} |h| |f_0^{-1}(h)| \leq 5 \sum_{h \in F(G)} |h| = 5 \cdot 2|E| = 5 \cdot 3|V|. \quad \blacksquare$$

Finalmente podemos concluir el resultado para grafos cúbicos planares sin puentes:

**Proposición 6.5.** *Toda secuencia creciente  $(G_k)_{k \in \mathbb{N}}$  de grafos planares cúbicos sin puentes satisface que  $\lim_{k \rightarrow \infty} cc(G_k) = 0$ .*

*Demostración.* Sea  $G_k = (V_k, E_k)$  un grafo de la secuencia y supongamos s.p.g. que  $|V_k| = k$ .

Luego

$$\begin{aligned}
 cc(G_k) &= \frac{1}{n} \sum_{v \in V_k} \frac{depth_{G_k}(v)}{ecc(v)} \\
 &\leq \frac{1}{2kr(G_k)} \sum_{v \in V_k} |f_0(v)| \\
 &\leq \frac{15}{2r(G_k)}.
 \end{aligned}$$

En donde usamos el hecho de que  $depth_{G_k}(v) = \left\lceil \frac{f_{min}(v)}{2} \right\rceil \leq f_0(v)$  y en donde  $f_0$  es la F-función de los lemas anteriores. Ahora, como  $r(G_k) = \Omega(\log(k))$ , se tendrá que  $cc(G_k) = O(\frac{1}{\log(k)})$ . Así concluimos que  $\lim_{k \rightarrow \infty} cc(G_k) = 0$ . ■

**Observación 1.** El resultado anterior puede ser extendido naturalmente a grafos planares regulares de grado 4 y sin puentes. Esto ya que como fue mencionado en 3.2, en esta clase de grafos los ciclos (y por ende las fronteras de caras) son también comunidades.

**Observación 2.** La condición de que los grafos no posean puentes es necesaria para el resultado anterior. De hecho, en la Figura 6.4 se exhibe un grafo cúbico planar de tamaño  $6n$  con  $cc(G) \geq \frac{n}{6(n+6)}$ .

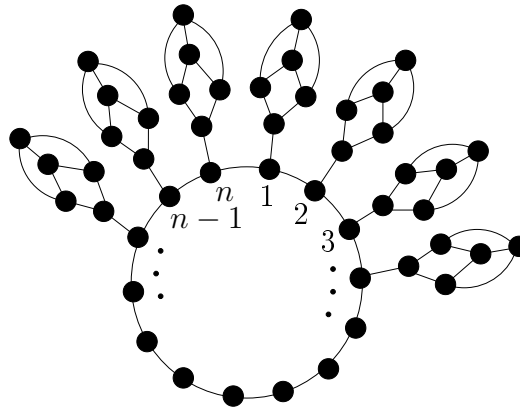


Figura 6.4: Un grafo cúbico planar con puentes.

# Bibliografía

- [1] Reinhard Diestel. *Graph Theory (Graduate Texts in Mathematics)*. Springer, August 2005.
- [2] M. Sipser. *Introduction to the Theory of Computation*. International Thomson Publishing, 1996.
- [3] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.
- [4] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang. Complex networks: Structure and dynamics. *Physics Reports*, 424(4-5):175–308, 2006.
- [5] P. Erdős and A. Rényi. On random graphs. *Publ. Math. Debrecen*, (6):290–297, 1959.
- [6] D. J. Watts. *Small Worlds: The Dynamics of Networks Between Order and Randomness*. Princeton University Press, 2003.
- [7] D. J. Watts and S. H. Strogatz. Collective dynamics of “small-world” networks. *Nature*, 393:440–442, 1998.
- [8] J. Kleinberg. The small-world phenomenon: An algorithmic perspective. In *Proc. 32nd ACM Symposium on Theory of Computing*, pages 151–161, 2000.
- [9] S.Ñ. Soffer and A. Vazquez. Network clustering coefficient without degree-correlation biases. *Phys. Rev. E*, 71(5):057101, 2005.
- [10] T. Schank and D. Wagner. Approximating clustering coefficient and transitivity. *Journal of Graph Algorithms and Applications*, 9(2):265–275, 2005.
- [11] A. H. Abdo and A. P. S. de Moura. Clustering as a measure of the local topology of networks. 2006. arXiv:physics/0605235.
- [12] G. Caldarelli, R. Pastor-Santorrás, and A. Vespignani. Cycles structure and local ordering in complex networks. *The European Physical Journal B - Condensed Matter*, 38(2):183–186, 2004.

## BIBLIOGRAFÍA

---

- [13] A. Fronczak, J. A. Holyst, M. Jedynek, and J. Sienkiewicz. Higher order clustering coefficients in Barabasi-Albert networks. *Physica A*, 316(1):688–694, 2002.
- [14] A.L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.
- [15] A.L. Barabási, R. Albert, and H. Jeong. Mean-field theory for scale-free random networks. *Physica A*, 272:173–187, 1999.
- [16] J. Buhl, J. Gautrais, R. V. Solé, P. Kuntz, S. Valverde, J.L. Deneubourg, and G. The-raulaz. Efficiency and robustness in ant networks of galleries. *The European Physical Journal B - Condensed Matter*, 42:123–129, 2004.
- [17] V. Latora and M. Marchiori. Efficient behavior of small-world networks. *Phys. Rev. Lett.*, 87(19):198701, 2001.
- [18] G. W. Flake, S. Lawrence, and C. L. Giles. Efficient identification of web communities. In *Proc. of the 6th ACM SIGKDD Int. Conference on Knowledge Discovery and Data Mining*, pages 150–160, 2000.
- [19] J. Bermond, J. Bond, D. Peleg, and S. Perennes. Tight bounds on the size of 2-monopolies. In *Proc. 3rd Colloq. on Structural Information and Communication Complexity*, 1996.
- [20] D. Peleg. Local majorities, coalitions and monopolies in graphs: A review. *Theor. Comput. Sci.*, 282(2):231–257, 2002.
- [21] D. Peleg. Size bounds for dynamic monopolies. In *Proc. of 4th Colloquium on Structural Information and Communication Complexity(SIROCCO 97)*, pages 151–161, 1997.
- [22] P. Flocchini, E. Lodi, F. Luccio, L. Pagli, and N. Santoro. Dynamic monopolies in tori. *Discrete Appl. Math.*, 137(2):197–212, 2004.
- [23] Paola Flocchini, Rastislav Královič, Peter Ružička, Alessandro Roncato, and Nicola Santoro. On time versus size for monotone dynamic monopolies in regular topologies. *J. of Discrete Algorithms*, 1(2):129–150, 2003.
- [24] M.E.J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69:026113, 2004.
- [25] M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E*, 74(3):036104, 2006.
- [26] G. W. Flake, R. E. Tarjan, and K. Tsioutsoulis. Graph clustering and minimum cut trees. *Internet Mathematics*, 1(4):385–408, 2004.

## BIBLIOGRAFÍA

---

- [27] H. Ino, M. Kudo, and A. Nakamura. Partitioning of web graphs by community topology. In *Proc. of the 14th International Conference on World Wide Web*, pages 661–669, 2005.
- [28] V. V. Vazirani. *Approximation Algorithms*. Springer, 2004.
- [29] B. V. Cherkassky, A. V. Goldberg, and T. Radzik. Shortest paths algorithms: theory and experimental evaluation. In *Proc. of the 5th Annual ACM-SIAM SODA*, pages 516–525, 1994.
- [30] Aric Hagberg, Dan Schult, and Pieter Swart. Networkx. high productivity software for complex networks. <https://networkx.lanl.gov/>.
- [31] A. J. Hoffman and R. R. Singleton. On Moore graphs with diameter 2 and 3. *IBM J. Res. Develop.*, (4):497–504, 1960.
- [32] M. Miller and J. Siran. Moore graphs and beyond: A survey of the degree/diameter problem. *Electronic Journal of Combinatorics*, (DS14), 2005.
- [33] Martin Loebel. Comunicación personal.