

# **Grafos Small-World**

Leonardo Torres



## Contents

Chapter 1. Introducción	5
Teoría de Grafos	5
Teoría de Redes	5
<i>Small-Worlds</i>	5
Chapter 2. Preliminares	7
Convenciones	7
Definiciones	8
Chapter 3. El modelo <i>small-world</i> de Cont y Tanimura	11
El modelo simplificado	11
Modelo generalizado	15
Appendix A. Algunos resultados	17



## CHAPTER 1

# Introducción

### Teoría de Grafos

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam et dignissim sem. Nullam maximus ante ac porttitor blandit. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Pellentesque quis quam tincidunt, suscipit sem quis, eleifend ipsum. Integer eget eleifend nisl. Maecenas et vulputate diam. Curabitur luctus consequat fringilla. Etiam pretium ligula et dui interdum, vitae porta erat maximus. Aliquam erat volutpat. Proin tempus nibh tincidunt quam eleifend venenatis.

### Teoría de Redes

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam et dignissim sem. Nullam maximus ante ac porttitor blandit. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Pellentesque quis quam tincidunt, suscipit sem quis, eleifend ipsum. Integer eget eleifend nisl. Maecenas et vulputate diam. Curabitur luctus consequat fringilla. Etiam pretium ligula et dui interdum, vitae porta erat maximus. Aliquam erat volutpat. Proin tempus nibh tincidunt quam eleifend venenatis.

### *Small-Worlds*

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam et dignissim sem. Nullam maximus ante ac porttitor blandit. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Pellentesque quis quam tincidunt, suscipit sem quis, eleifend ipsum. Integer eget eleifend nisl. Maecenas et vulputate diam. Curabitur luctus consequat fringilla. Etiam pretium ligula et dui interdum, vitae porta erat maximus. Aliquam erat volutpat. Proin tempus nibh tincidunt quam eleifend venenatis.

**Definiciones (énfasis en plural).** Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam et dignissim sem. Nullam maximus ante ac porttitor blandit. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Pellentesque quis quam tincidunt, suscipit sem quis, eleifend ipsum. Integer eget eleifend nisl. Maecenas et vulputate diam. Curabitur luctus consequat fringilla. Etiam pretium ligula et dui interdum, vitae porta erat maximus. Aliquam erat volutpat. Proin tempus nibh tincidunt quam eleifend venenatis.

**Algunos resultados.** Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam et dignissim sem. Nullam maximus ante ac porttitor blandit. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Pellentesque quis quam tincidunt, suscipit sem quis, eleifend ipsum. Integer eget

eleifend nisl. Maecenas et vulputate diam. Curabitur luctus consequat fringilla. Etiam pretium ligula et dui interdum, vitae porta erat maximus. Aliquam erat volutpat. Proin tempus nibh tincidunt quam eleifend venenatis.

**En el presente trabajo,...** Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam et dignissim sem. Nullam maximus ante ac porttitor blandit. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Pellentesque quis quam tincidunt, suscipit sem quis, eleifend ipsum. Integer eget eleifend nisl. Maecenas et vulputate diam. Curabitur luctus consequat fringilla. Etiam pretium ligula et dui interdum, vitae porta erat maximus. Aliquam erat volutpat. Proin tempus nibh tincidunt quam eleifend venenatis.

## CHAPTER 2

# Preliminares

### Convenciones

Antes de comenzar, indicaremos las convenciones de notación que usaremos en el resto de nuestra discusión.

Si  $A$  es cualquier conjunto, escribiremos  $|A|$  para denotar su cardinalidad. Cada vez que hablemos de un grafo  $G = (\mathcal{N}, \mathcal{E})$ , identificaremos los nodos  $i \in \mathcal{N}$  con los números naturales  $1, 2, \dots, |N|$ . Más aún, identificaremos a  $G$  con el conjunto de nodos,  $G \equiv \mathcal{N}$ , es decir  $i \in G \iff i \in \mathcal{N}$ ,  $\forall i = 1, 2, \dots, |G|$  y  $|G| = |\mathcal{N}|$ . Por otro lado, escribiremos la arista dirigida que va de  $i$  a  $j$  como el par ordenado  $(i, j)$ . En nuestro caso, como trataremos siempre con aristas no dirigidas, tendremos que  $(i, j) \in \mathcal{E}$  implica  $(j, i) \in \mathcal{E}$ . Por ello, escribiremos  $(i, j)$  también para la arista no dirigida que une  $i$  con  $j$ . En este caso, diremos que  $i$  y  $j$  son *vecinos* y escribiremos  $V(i)$  para el conjunto de nodos vecinos de  $i$ . Para complicar las cosas, a veces identificaremos también a  $G$  con su conjunto de aristas  $\mathcal{E}$ . Esto no causará ambigüedad, ya que los nodos son números, mientras que las aristas son pares ordenados. De esta manera, escribiremos  $i \in G$  y  $(i, j) \in G$  sin lugar a confusión. Mientras que escribimos  $(i, j) \in G$  para las aristas *existentes* en el grafo  $G$ , escribiremos  $(i, j) \in G \times G$  para una arista *posible* del grafo. Por ejemplo, si  $G = (\{1, 2, 3\}, \{(1, 2), (2, 1)\})$ , son verdaderas las afirmaciones  $|G| = 3$ ,  $(1, 2) \in G$  y  $(1, 3) \in G \times G$ .

A continuación, definiremos los espacios con los que trabajaremos. Llamaremos  $\Gamma_n$  al conjunto de grafos simples, no dirigidos, de aristas binarias y de  $n$  nodos,

$$\Gamma_n = \{G = (\mathcal{N}, \mathcal{E}) : |G| = n, (i, i) \notin G, (i, j) \in G \iff (j, i) \in G\}.$$

Usualmente, identificaremos cada grafo con su matriz de adyacencia. De ahí que el conjunto  $\Gamma_n$  esté en biyección con el conjunto de matrices simétricas con entradas 0 ó 1: si  $G \in \Gamma_n$  entonces  $G \equiv M$ ,  $M \in \{0, 1\}^{n \times n}$ , donde  $M$  es la matriz de adyacencia correspondiente a  $G$ . Por tratarse de grafos simples, si las entradas de la matriz de adyacencia  $M$  son  $(a_{ij})$ , tendremos  $a_{ii} = 0$ ,  $\forall i = 1, 2, \dots, |G|$ . Además, escribiremos  $\Gamma_\infty = \bigcup_{n=1}^{\infty} \Gamma_n$  para el conjunto de todos los grafos simples, no dirigidos y de aristas binarias.

A este último conjunto lo dotaremos de un  $\sigma$ -álgebra, convirtiéndolo en un espacio medible. En realidad, usaremos el  $\sigma$ -álgebra usual de boreleanos del conjunto de matrices  $\{0, 1\}^{n \times n}$ , y lo trasladaremos a  $\Gamma_n$  mediante la biyección mencionada. Bajo este  $\sigma$ -álgebra, llamémosle  $\mathcal{B}$ , una función  $f : \Gamma_\infty \rightarrow \mathbb{R}$  es  $\mathcal{B}$ -medible si cada una de sus proyecciones  $f|_n : \Gamma_n \rightarrow \mathbb{R}$  es medible, cuando es vista como función del espacio de matrices.

### Definiciones

Pasaremos ahora a definir los elementos sobre los que construiremos los modelos que queremos estudiar.

**Indicadores gráficos.** La función  $f : \Gamma_\infty \rightarrow \overline{\mathbb{R}}$  es un indicador gráfico si es  $\mathcal{B}$ -medible. Por lo mencionado anteriormente, para ello es suficiente expresarla como una función medible de las matrices de adyacencia. En nuestra discusión, nos interesarán en particular cuatro indicadores. Pasamos a definir tres de ellos y dejaremos el último para una sección posterior. Sea el grafo  $G \in \Gamma_\infty$ , con matriz de adyacencia  $M = (a_{ij})$  y consideremos los nodos  $i$  y  $j$  de  $G$ . Entonces definimos los siguientes indicadores.

*Grado y grado promedio.* Estos indicadores son las cantidades típicas que se manejan en todas las áreas de la Teoría de Grafos.  $\deg_G(i)$  es el número de vecinos del nodo  $i$ , mientras que  $\deg(G)$  es la media aritmética de ellos.

$$\deg_G(i) = \sum_{j=1}^{|G|} a_{ij} \quad y \quad \deg(G) = \frac{1}{|G|} \sum_{i=1}^{|G|} \deg_G(i).$$

*Coefficiente de agrupación local.* Hay varios coeficientes de agrupación usados en Teoría de Redes, todos los cuales miden el grado de transitividad de un grafo de maneras ligeramente distintas. El que nos interesará a nosotros es el *coeficiente de agrupación local*, definido de la siguiente manera,

$$c_i = \frac{\sum_{j=k \neq i}^{|G|} a_{ij} a_{ik} a_{jk}}{\deg_G(i) \binom{\deg_G(i)-1}{2}}.$$

Esta expresión puede parecer atemorizante a primer vista. Sin embargo, ella se lee como sigue. *El coeficiente de agrupación local del nodo  $i$  es el número de triángulos presentes con un vértice en  $i$ , sobre el número total de triángulos que  $i$  podría formar.* Como las entradas  $a_{ij}$  son solo 0 ó 1, y como los nodos  $i, j, k$  forman un triángulo solo si el producto  $a_{ij} a_{jk} a_{ik}$  es igual a 1, el numerador de la expresión es igual al número de triángulos presentes en el grafo con un vértice en  $i$ . Por otro lado, el denominador es simplemente el número total de posibles triángulos que incluyen a  $i$ .

*Diámetro.* Es un indicador global del “tamaño” del grafo.

$$\text{diam}(G) = \max_{i \in G, j \in G} \{d(i, j)\},$$

donde  $d(i, j)$  es el menor número de aristas en todos los caminos que empiezan en  $i$  y terminan en  $j$ .

Claramente, estas funciones tienen el derecho de llamarse indicadores gráficos, desde que las podemos expresar como funciones medibles de las matrices de adyacencia.

Definidos nuestro espacio medible y funciones medibles sobre él, pasaremos a incluir aleatoriedad en nuestro estudio. En lo subsiguiente, supondremos que existe un espacio de probabilidad abstracto  $(\Omega, \mathcal{F}, \mathcal{P})$  que contiene el espacio muestral de todos los resultados posibles de nuestros experimentos.



**Grafos e indicadores aleatorios.** Un grafo aleatorio es una variable aleatoria  $G : \Omega \longrightarrow \Gamma_\infty$ . Dados un grafo aleatorio  $G$  y un indicador gráfico  $Q$ , la variable aleatoria  $Q \circ G : \Omega \longrightarrow \mathbb{R}$  se llama indicador aleatorio. El cuarto indicador que faltaba mencionar es un indicador que *siempre* es aleatorio, sin importar si el grafo que mide lo es o no.

*Distancia típica.* Dado un grafo  $G \in \Gamma_\infty$ , aleatorio o determinístico, definimos

$$T(G) = d(u, v)$$

donde la arista  $(u, v)$  es escogida de manera aleatoria y uniforme de  $G \times G \setminus \{(i, i) : i \in G\}$ .

**Modelo de crecimiento de grafos.** Un modelo de crecimiento será una sucesión  $(G_n)_{n=1}^\infty$  donde  $G_n \in \Gamma_n$ . Si todos los  $G_n$  son grafos aleatorios, la sucesión será un modelo de crecimiento aleatorio.

**Comportamiento asintótico de indicadores.** Para este trabajo, daremos un modelo específico de crecimiento aleatorio y nos interesará estudiar el comportamiento de ciertos indicadores cuando  $|G| \rightarrow \infty$ . Sean  $(G_n)_{n=1}^\infty$  un modelo de crecimiento (determinístico) y  $Q$  un indicador. Si  $f : \mathbb{N} \longrightarrow \mathbb{R}$  es una sucesión determinística de números, decimos que es una

*Cota superior de crecimiento de  $(G_n)$ .* Si se cumple

$$\limsup_{n \rightarrow \infty} \frac{Q(G_n)}{f(n)} \leq 1.$$

O una

*Cota inferior de crecimiento de  $(G_n)$ .* Si

$$\limsup_{n \rightarrow \infty} \frac{Q(G_n)}{f(n)} \geq 1.$$

En el caso en que  $f(n)$  sea cota superior y otra sucesión  $g(n)$  sea cota inferior del crecimiento de  $(G_n)$  y si, además, se cumple

$$\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = c \in \mathbb{R},$$

diremos que  $Q(G_n)$  crece como  $f(n)$ .

**Comportamiento asintótico de indicadores aleatorios.** Ahora, consideremos un modelo de crecimiento aleatorio  $(G_n)_{n=1}^\infty$ . Sean  $Q$  un indicador y  $f$  una sucesión de números. En este caso,  $f$  será una

*Cota superior del crecimiento de  $(G_n)$  en esperanza.* Si

$$\limsup_{n \rightarrow \infty} \frac{\mathbb{E}[Q(G_n)]}{f(n)} \leq 1.$$

*Cota superior del crecimiento de  $(G_n)$  en probabilidad.* Cuando

$$\limsup_{n \rightarrow \infty} P\left(\frac{Q(G_n)}{f(n)} \leq 1\right) = 1.$$

*Cota superior del crecimiento de  $(G_n)$  casi ciertamente.* En el caso

$$P\left(\limsup_{n \rightarrow \infty} \frac{Q(G_n)}{f(n)} \leq 1\right) = 1.$$

Análogamente, se definen las cotas inferiores en esperanza, en probabilidad y casi ciertas. Claramente, estas definiciones van en orden ascendente de rigor.

**Small-World.** Finalmente, podemos presentar la definición formal de grafo *small-world* usada por Cont y Tanimura.

Un modelo de crecimiento aleatorio  $(G_n)_{n=1}^\infty$  se llama *small-world* si

- (1)  $\exists a_1 \geq 0 : f(n) = a_1 \log(n)$  es cota superior del crecimiento de  $\deg(G_n)$ ,
- (2)  $\exists a_2 > 0 : c_i \geq a_2, \forall i \in G_n \forall n = 1, 2, \dots, \infty$ ,
- (3)  $\exists a_3 \geq 0 : g(n) = a_3 \log(n)$  es cota superior del crecimiento de  $T(G_n)$ .

Si el modelo  $(G_n)$  es aleatorio, cualquiera de las tres propiedades puede cumplirse en esperanza, en probabilidad o casi ciertamente.

## CHAPTER 3

### El modelo *small-world* de Cont y Tanimura

#### El modelo simplificado

Este modelo nace del deseo de construir un modelo de grafo *small-world* que no dependa de una estructura regular subyacente. La idea principal es empezar con un número de “comunidades” (subgrafos completos) y unirlos de manera aleatoria. En esta versión simplificada, consideraremos el caso en que todas las comunidades tienen el mismo número de nodos. Esta restricción es para dejar más clara la prueba y familiarizarnos con el modelo. En la siguiente sección, levantaremos esta restricción.

*Construcción.* Fijemos los números  $n, \delta, M \in \mathbb{N}$  con  $n = \delta \times M$  y consideremos  $G \in \Gamma_n$  el grafo de  $n$  nodos sin aristas. Añadiremos aristas a  $G$  en tres pasos.

- (1) Particionamos los nodos  $1, 2, \dots, n$  en  $M$  conjuntos disjuntos,  $G^1, G^2, \dots, G^M$ , con  $\delta$  nodos cada uno. Si  $i \in G^k$ , decimos que  $G^k$  es la *comunidad de origen del nodo  $i$* . Cada  $i$  lo unimos con una arista a cada nodo que tenga la misma comunidad de origen:  $i, j \in G^k \implies (i, j) \in G$ .
- (2) Para cada  $i \in G$ , sean las variables aleatorias  $X_i : \Omega \longrightarrow \{G^1, G^2, \dots, G^M\}$ , independientes e idénticamente distribuidas de manera uniforme sobre el conjunto de comunidades,  $X_i \sim U(G^1, G^2, \dots, G^M)$ . En la práctica, identificaremos cada  $G^k$  con el número  $k$ , excepto en casos ambiguos. De esta manera, si  $X_i = k$ , diremos que  $G^k$  es la *comunidad secundaria o de destino del nodo  $i$* . A veces diremos que el nodo  $i$  *escoge*  $X_k$  como destino. Para cada  $i$ , añadimos a  $G$  las aristas  $(i, j)$ ,  $\forall j \in X_i$ . Es decir, unimos cada nodo con todos los nodos en su comunidad secundaria.
- (3) Si los nodos  $i, j$  cumplen  $X_i = X_j$ , añadimos a  $G$  la arista  $(i, j)$ . Esto es, unimos todos los nodos que tengan la misma comunidad secundaria.

Cabe recordar que, como  $G$  es un grafo no dirigido, basta añadir la arista  $(i, j)$  para que  $(j, i)$  también esté presente.

Podemos describir la construcción de  $G$  de la siguiente manera. Dado  $m = 1, 2, \dots, M$ , sea

$$A^m = G^m \cup \{i : X_i = G^m\}$$

el conjunto de los nodos que tienen  $G^m$  como comunidad de origen o de destino. Entonces,

$$(i, j) \in G \iff i \neq j \text{ and } (i, j) \in \bigcup_{m=1}^M A^m \times A^m.$$

De esta manera, podemos ver a  $G$  como los  $M$  subgrafos completos  $A^m$  que están unidos de manera aleatoria.

Hecha la misma construcción para cada  $n$  (o, mejor dicho, para cada  $n$  múltiplo de  $\delta$ ), queremos demostrar la siguiente afirmación.

AFIRMACIÓN 1. El modelo  $(G_n)_{n=1}^\infty$  es un grafo *small-world*.

La demostración la dividimos en tres partes, una para cada característica propia de los grafos *small-world*. Los resultados propios del grado y el coeficiente de agrupación local corresponden a las Proposiciones 1 y 2, respectivamente. El estudio de la distancia típica abarca las Proposiciones 3 y 4.

PROPOSICIÓN 1. *El grado esperado del modelo está acotado superiormente por  $4\delta - 1$ .*

PROOF. Escribimos  $S_m = |\{i : X_i = m\}|$ , de donde  $|A^m| \leq \delta + S_m$ . El número de aristas en  $G$  está acotado por

$$|\mathcal{E}| \leq \sum_{m=1}^M \frac{|A^m|(|A^m| - 1)}{2} \leq \sum_{m=1}^M \frac{(\delta + S_m)(\delta + S_m - 1)}{2}.$$

Por ser las variables  $X_i$  i.i.d. y uniformes, tenemos

$$P(S_1 = x_1, S_2 = x_2, \dots, S_M = x_M) = \begin{cases} \frac{n!}{x_1!x_2!\dots x_M!} \left(\frac{1}{M}\right)^n, & \sum_{k=1}^M x_k = n, \\ 0, & \text{otherwise} \end{cases},$$

i.e., el vector  $(S_1, S_2, \dots, S_M)$  tiene distribución multinomial. Entonces, tenemos

$$\mathbb{E}[S_m] = \frac{n}{M} = \delta, \quad \text{Var}[S_m] = \frac{n}{M} \left(1 - \frac{1}{M}\right) = \delta \left(1 - \frac{1}{M}\right), \quad \mathbb{E}[S_m^2] = \delta \left(1 - \frac{1}{M}\right) + \delta^2$$

Y, en consecuencia,

$$\mathbb{E}[|\mathcal{E}|] \leq \frac{1}{2} \sum_{m=1}^M \mathbb{E}[(\delta + S_m)(\delta + S_m - 1)] = \frac{\delta}{2} (4\delta M - M - 1).$$

Finalmente,

$$\mathbb{E}[\deg(G)] = \frac{2\mathbb{E}[|\mathcal{E}|]}{n} = 4\delta - 1 - \frac{\delta}{n}.$$

□

PROPOSICIÓN 2. *En el modelo  $(G_n)_{n=1}^\infty$ , el coeficiente de agrupación local cumple*

$$c_i \geq \frac{1}{2} - \frac{1}{\delta - 1}, \quad \forall i \in G_n, \forall n = 1, 2, \dots$$

PROOF. Fijado un nodo  $i$ , sean  $G^j$  y  $G^k$  sus comunidades de origen y destino, respectivamente. Entonces, podemos particionar  $G(i)$  de la siguiente manera:

$$G(i) = (G^j \cup \{u : X_u = G^j\}) \setminus \{i\} \bigcup (G^k \cup \{u : X_u = G^k\}) \setminus \{i\}.$$

Ambos conjuntos tienen al menos  $\delta - 1$  nodos. Aplicando el Lema 1 del Apéndice A, con  $l = 2$ ,  $h \geq \delta - 1$ , tenemos

$$c_i \geq \frac{1}{2} - \frac{1}{\delta - 1}.$$

□

OBSERVACIÓN. Para que el modelo sea *small-world*, necesitamos que cada  $c_i > 0$ . Claramente, esto ocurrirá en el modelo simplificado cuando  $\delta \geq 3$ .

PROPOSICIÓN 3. *En el modelo  $(G_n)_n^\infty$ , cuando  $n \rightarrow \infty$ , es cada vez más probable que podamos alcanzar por lo menos la mitad de los nodos en un número logarítmico de pasos, habiendo empezado desde un nodo arbitrario.*

*Concretamente, para  $n$  suficientemente grande, sea un nodo  $u \in G_n$ . Entonces, existe un conjunto de nodos  $\bar{S}(u, n) \subseteq G_n$  tal que  $|\bar{S}| \geq \frac{n}{2}$  y además, si  $v \in \bar{S}$ , entonces*

$$P(d(u, v) \leq 4 \log(n) + 4) \geq 1 - \frac{1}{\delta^2 n^2}.$$

PROOF. Fijamos  $n$  para el resto de nuestro argumento. Probaremos que, si empezamos desde una comunidad arbitraria, digamos  $G^1$ , podemos alcanzar al menos la mitad de las demás comunidades. Como cada comunidad es un subgrafo completo y todas ellas tienen el mismo número  $\delta$  de nodos, esto probará la Proposición. En lo subsiguiente, identificaremos el conjunto de comunidades  $\{G^1, G^2, \dots, G^M\}$  con el conjunto de números  $\{1, 2, \dots, M\}$ .

Definimos el proceso de descubrimiento de comunidades  $C_k, N_k$ , empezando desde  $G^1 \equiv 1$ , recursivamente.

$$C_0 = \{1\}, \quad N_k = \{i \in G_n : \exists d \in C_k, i \in d\} = \bigcup_{d \in C_k} G^d$$

$$C_k = \{G \notin \bigcup_{l=0}^{k-1} C_l : \exists j \in N_{k-1}, X_j = G\}, \quad k = 0, 1, \dots$$

En palabras,  $C_k$  es el conjunto de comunidades que se pueden alcanzar en, como mucho,  $k$  pasos; mientras que  $N_k$  es el conjunto de nodos de estas comunidades. Notamos que los  $C_k$  son disjuntos dos a dos.

También definimos los eventos

$$E_k = \{|C_k| \geq 3|C_{k-1}|\}, \quad k = 1, 2, \dots$$

$E_k$  implica que el número de comunidades que alcanzamos en el último paso es al menos tres veces mayor que el número anterior.  $\bigcap_{m=1}^l E_m$  implica que hemos descubierto comunidades a una velocidad exponencial, con base 3, hasta el  $l$ -ésimo paso.

Ahora, estamos interesados en estimar la probabilidad de descubrir nuevas comunidades a una velocidad exponencial hasta descubrir al menos la mitad de ellas. Para ello, definimos los tiempos de parada

$$\tau_u = \min\{t \in \mathbb{N} : \sum_{j=0}^t |C_j| \geq u\}, \quad u = 1, 2, \dots$$

Para un  $T$  dado, estimamos la probabilidad  $P(\bigcap_{j=1}^T E_k) = \prod_{j=1}^T P(E_k | \bigcap_{m=1}^{j-1} E_m)$ .<sup>1</sup>

Estudiamos el lado derecho de la igualdad, para cada  $j$ , según tres casos distintos:  $j = 1$ ,  $2 \leq j \leq \tau_{\sqrt{M}}$  y  $\tau_{\sqrt{M}} < j \leq \tau_{\frac{M}{2}}$ . Queremos encontrar que estas probabilidades son cada vez más grandes.

<sup>1</sup>Donde tomamos la convención  $P(E_1 | \bigcap_{m=1}^{m=0} E_m) = P(E_1)$ .

$j = 1$ : Estimamos la probabilidad  $P(E_1^c) = P(|C_1| \leq 2)$ .  $\{|C_1| \leq 2\}$  ocurre cuando todos los nodos de  $G^1$  eligieron como comunidades secundarias a  $\{1, k_1, k_2\}$ , para algún par  $k_1, k_2$ . Tenemos

$$\begin{aligned} \{|C_1| \leq 2\} &= \bigcup_{k_1, k_2 \in \{1, 2, \dots, M\}, k_1 \neq k_2} \{X_i \in \{1, k_1, k_2\}, \forall i \in G^1\} \\ P(|C_1| \leq 2) &\leq \sum_{k_1 \neq k_2} P(X_i \in \{1, k_1, k_2\}, \forall i \in G^1) \\ P(|C_1| \leq 2) &\leq \sum_{k_1 \neq k_2} \prod_{i=1}^{|G^1|} P(X_i \in \{1, k_1, k_2\}) = \binom{M}{2} \left(\frac{3}{M}\right)^\delta = \frac{3^\delta}{2} \frac{M-1}{M^{\delta-1}} \leq \frac{3^\delta}{M^{\delta-2}}. \end{aligned}$$

Escogemos  $\delta \geq 6$  y  $M$  suficientemente grande tal que  $\frac{3^\delta}{M^{\delta-2}} \leq \frac{1}{M^3}$ .

$2 \leq j \leq \tau_{\sqrt{M}}$ : **POR PROBAR**

$\tau_{\sqrt{M}} < j \leq \tau_{\frac{M}{2}}$ : **POR PROBAR**

Conseguida esta cota, pasamos a acotar  $P(\tau_{\frac{M}{2}} \geq \log(M))$ . Queremos que esta probabilidad sea cada vez más pequeña.

Definamos  $L = \min\{t \in \mathbb{N} : e^t \geq \frac{M}{2}\} < \log(M) + 1$  y el tiempo de parada acotado  $T = \min\{L, \tau_{\frac{M}{2}}\}$ . Tenemos que

- el evento  $\bigcap_{m=1}^L E_m$  implica  $|C_L| > e^L \geq \frac{M}{2}$ ,
- el evento  $\{j \leq \tau_{\frac{M}{2}}\} \cap \bigcap_{m=1}^j E_m$  implica  $|C_{j-1}| < \frac{M}{2}$ ,
- el evento  $\bigcap_{m=1}^T E_m$  implica  $L \geq \tau_{\frac{M}{2}}$ .

Con todo, tenemos

$$P(L \geq \tau_{\frac{M}{2}}) \geq P\left(\bigcap_{m=1}^T E_m\right) = \prod_{j=1}^T P(E_k | \bigcap_{m=1}^{j-1} E_m) \geq \left(1 - \frac{1}{M^3}\right)^T \geq \left(1 - \frac{1}{M^3}\right)^{\log(M)+1}$$

Para concluir, usamos el Lema 2 del Apéndice A, con  $f(M) = M^3$ ,  $g(M) = \log(M)$ . Tenemos, para  $M$  suficientemente grande,

$$P(\tau_{\frac{M}{2}} < \log(M) + 1) \leq P(\tau_{\frac{M}{2}} > L) \leq 1 - \left(1 - \frac{1}{M^3}\right)^{\log(M)+1} \leq \frac{\log(M) + 1}{M^3}.$$

Finalmente, definimos el conjunto de comunidades  $S = \bigcup_{j=0}^{\tau_{\frac{M}{2}}} C_j$ , que cumple  $|S| \geq \frac{M}{2}$ , por definición de  $\tau_{\frac{M}{2}}$ . Ahora bien,  $\tau_{\frac{M}{2}} < \log(M) + 1$  implica

$$\forall G, G' \in S : d(G, G') \leq d(G, S) + d(S, G') \leq 2\log(M) + 2.$$

Por lo tanto,

$$P(d(G, G') \leq 2\log(M) + 2) \geq P(\tau_{\frac{M}{2}} \leq \log(M) + 1) \geq 1 - \frac{\log(M) + 1}{M^3}.$$

Basta tomar los nodos  $\bar{S} = \bigcup_{G \in S} G$ , un  $M = n \times \delta$  grande para el cual  $\frac{\log(M)+1}{M^3} \leq \frac{1}{M^2}$  y observar que si  $u \in G, v \in G'$ , entonces  $d(u, v) \leq 2d(G, G')$ , para obtener

$$P(d(u, v) \leq 4\log(n) + 4) \geq 1 - \frac{1}{\delta^2 n^2}, \quad \forall u, v \in \bar{S}$$

□

PROPOSICIÓN 4. *La distancia típica en  $(G_n)_{n=1}^\infty$  cumple*

$$P(\limsup_{n \rightarrow \infty} \frac{T(G_n)}{O(\log(n))} \leq 1) = 1$$

*donde  $O(\log(n))$  es una función lineal de  $\log(n)$ .*

PROOF. Sean  $u, v$  nodos elegidos de manera uniforme en  $G \times G \setminus \{(i, i) : i \in G\}$ ,

$G^u$  y  $G^v$  sus comunidades de origen, y sean  $S_u = \bigcup_{j=1}^{\tau_{\frac{M}{2}}^u} C_j^u$ ,  $S_v = \bigcup_{j=1}^{\tau_{\frac{M}{2}}^v} C_j^v$ , donde  $C_j^u$ ,  $C_j^v$  son los conjuntos de todas las comunidades que se pueden alcanzar en, como mucho,  $j$  pasos, a partir desde  $G^u$ ,  $G^v$ , respectivamente. Ahora bien,  $S_u \cap S_v \neq \emptyset$  pues  $|S_u| + |S_v| \geq M$ . Entonces, sea  $l \in S_u \cap S_v$ . Tenemos

$$d(u, v) \leq d(u, l) + d(l, v).$$

De la proposición anterior, tenemos que, si  $n$  es suficientemente grande, entonces  $P(d(u, v) > 8 \log(n) + 8) \leq \frac{2}{\delta^2 n^2}$ . En consecuencia,

$$\sum_{n=1}^{\infty} P(\frac{d(u, v)}{O(\log(n))} > 1) < \infty.$$

Donde  $O(\log(n)) = 8 \log(n) + 8$  es una función lineal de  $\log(n)$ .

Por el Lema de Borel-Cantelli, tenemos

$$P(\limsup_{n \rightarrow \infty} \frac{d(u, v)}{O(\log(n))} > 1) = 0.$$

Es decir,

$$P(\limsup_{n \rightarrow \infty} \frac{T(G_n)}{O(\log(n))} \leq 1) = 1.$$

□

Con esto, hemos probado la Afirmación 1, con lo que el modelo es *small-world*. Sin embargo, Cont y Tanimura mejoran su resultado, con la siguiente proposición.

PROPOSICIÓN 5. *El diámetro cumple*

$$P(\limsup_{n \rightarrow \infty} \frac{\text{diam}(G_n)}{O(\log(n))} \leq 1) = 1.$$

PROOF. **POR PROBAR**

□

### Modelo generalizado

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam et dignissim sem. Nullam maximus ante ac porttitor blandit. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Pellentesque quis quam tincidunt, suscipit sem quis, eleifend ipsum. Integer eget eleifend nisl. Maecenas et vulputate diam. Curabitur luctus consequat fringilla. Etiam pretium ligula et dui interdum, vitae porta erat maximus. Aliquam erat volutpat. Proin tempus nibh tincidunt quam eleifend venenatis.





## APPENDIX A

### Algunos resultados

LEMA 1. Sea  $G \in \Gamma_n$  un grafo arbitrario y sean  $i \in G$  un nodo y  $G(i) = G \cap (V(i) \times V(i))$  el subgrafo de  $G$  que contiene todos los vecinos de  $i$ , pero no  $i$ , y todas las aristas entre ellos. Supongamos que  $G(i)$  se puede particionar en  $l$  subgrafos completos disjuntos. Entonces, se cumple

$$c_i \geq \frac{1}{l} - \frac{1}{h},$$

donde  $h$  es el número de nodos en el subgrafo más pequeño de  $G(i)$ .

PROOF. Consideremos tal partición y pongamos  $q_1, q_2, \dots, q_l$  para el número de nodos en cada subgrafo completo de  $G(i)$ . Tenemos  $a = \sum_{j=1}^l \frac{q_j(q_j-1)}{2}$  aristas en  $G(i)$ ,

mientras que el número total de posibles aristas es  $t = \frac{(\sum_{j=1}^l q_j)(-1 + \sum_{j=1}^l q_j)}{2}$ .

Luego,

$$c_i \geq \frac{a}{t} \geq \frac{\sum_j q_j(q_j-1)}{(\sum_j q_j)^2} \geq \frac{\overbrace{\sum_j q_j^2}^{\mathcal{A}}}{(\sum_j q_j)^2} - \frac{\overbrace{\sum_j q_j}^{\mathcal{B}}}{\sum_j q_j^2}.$$

Tenemos para  $\mathcal{A}$ , usando  $2q_j q_k \leq q_k^2 + q_j^2$ ,

$$(\sum_j q_j)^2 \leq \sum_j q_j^2 + \frac{1}{2} \sum_{k \neq j}^l (q_k^2 + q_j^2) = \sum_j q_j^2 + \frac{1}{2} \sum_{k \neq j}^l q_k^2 + \frac{1}{2} \sum_{k \neq j}^l q_j^2$$

$$(\sum_j q_j)^2 \leq \sum_j q_j^2 + (l-1) \sum_j q_j^2 = l \sum_j q_j^2.$$

Es decir,  $\mathcal{A} \geq \frac{1}{l}$ .

Por otro lado, notamos que  $\sum_j h q_j \leq \sum_j q_j^2$ , ó  $\mathcal{B} \leq \frac{1}{h}$ .

Finalmente,  $c_i \geq \frac{1}{l} - \frac{1}{h}$ . □

LEMA 2. Sean  $f(n)$ ,  $g(n)$  dos funciones tal que  $\lim_{n \rightarrow \infty} f(n) = \lim_{n \rightarrow \infty} g(n) = \infty$ , donde  $g(n)$  es no decreciente. Supongamos que  $\lim_{n \rightarrow \infty} h(n) = \frac{f(n)}{g(n)} = 0$ . Entonces, existe un  $n_0$  tal que

$$\forall n \geq n_0 : (1 - \frac{1}{f(n)})^{g(n)} \geq 1 - \frac{g(n)}{f(n)}.$$

PROOF. Dado,  $i \in \mathbb{N}$ , tenemos  $g(n)^i \geq g(n)$  y además,

$$\frac{-g(n)}{i f(n)^i} \geq \frac{-g(n)^i}{i f(n)^i}$$

$$g(n) \sum_{i=1}^{\infty} \frac{(-1)^{i+1}}{i} \left(\frac{-1}{f(n)}\right)^i \geq \sum_{i=1}^{\infty} \frac{(-1)^{i+1}}{i} \left(\frac{-g(n)}{f(n)}\right)^i.$$

Para algún  $n_0$ , tendremos  $f(n) > g(n) > 1$ . Luego,

$$g(n) \log\left(1 - \frac{1}{f(n)}\right) \geq \log\left(1 - \frac{g(n)}{f(n)}\right)$$

$$\left(1 - \frac{1}{f(n)}\right)^{g(n)} \geq 1 - \frac{g(n)}{f(n)}$$

□

LEMA 3. *Derivación de la media y varianza de la distribución multinomial.*

PROOF. **POR PROBAR**

□

LEMA 4. *Borel-Cantelli.*

PROOF. **POR PROBAR**

□