

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

---

Scuola di Scienze  
Corso di Laurea in Matematica

# Grafi aleatori: il modello di Erdős-Rényi

Tesi di Laurea in Geometria

Relatore:  
Chiar.mo Prof.  
Massimo Ferri

Presentata da:  
Anna Vesely

Sessione III  
Anno Accademico 2014-2015



*A Lucio, Luca e Claudio,  
che mi sono stati accanto  
quando più ne avevo bisogno*



# Introduzione

La teoria dei grafi aleatori, sviluppata per la prima volta nel 1959 dai matematici ungheresi Paul Erdős e Alfréd Rényi e, indipendentemente, dal matematico americano Edgar Gilbert, è un punto di incontro tra la teoria dei grafi e quella della probabilità.

In tale teoria si considerano grafi generati in maniera casuale secondo una determinata distribuzione di probabilità e, tramite un approccio probabilistico, se ne studiano le proprietà e il comportamento asintotico quando il numero di nodi tende all'infinito.

Qui verrà analizzato il modello sviluppato da Erdős e Rényi (modello ER), che descrive un particolare processo di generazione casuale di grafi, con lo scopo di studiare la connessione dei grafi ER ed esplicitare le loro possibili configurazioni al variare dei parametri.

Inoltre, si darà un esempio di applicazione per il modello.

In particolare, nel capitolo 1 sarà introdotta la nozione di grafo insieme ad alcune caratteristiche che ne rappresentino la connessione a livello locale e globale (grado e clustering), nonché il concetto di proprietà monotona.

Nel capitolo 2 verrà data una definizione formale del modello ER, per poi esaminare le caratteristiche precedentemente introdotte in questo caso particolare. Inoltre, si studieranno le configurazioni di un grafo ER in dipendenza dalla probabilità con cui esso è generato.

Nel capitolo 3, infine, si considererà una possibile applicazione del modello,

che sarà utilizzato per descrivere il processo di diffusione di una malattia infettiva.

# Indice

<b>Introduzione</b>	<b>i</b>
<b>1 Nozioni preliminari</b>	<b>1</b>
1.1 Notazione . . . . .	1
1.2 Grafi . . . . .	1
1.3 Connessione . . . . .	2
1.4 Clustering . . . . .	4
1.5 Proprietà . . . . .	6
<b>2 Il modello ER</b>	<b>7</b>
2.1 Grafo di Erdős-Rényi . . . . .	7
2.2 Grado dei nodi . . . . .	8
2.2.1 Caso generale . . . . .	8
2.2.2 Grafo con $n$ grande e $p$ piccolo . . . . .	10
2.3 Clustering . . . . .	12
2.4 Componente gigante . . . . .	13
2.5 Proprietà e funzioni soglia . . . . .	15
2.6 Configurazioni di un grafo ER . . . . .	17
<b>3 Esempio di applicazione: diffusione</b>	<b>21</b>
3.1 Diffusione su grafo ER . . . . .	21
3.2 Modello SIS . . . . .	23
3.2.1 Costruzione del modello . . . . .	24
3.2.2 Fase stazionaria . . . . .	26

3.2.3	Modello SIS su grafo ER . . . . .	30
<b>4</b>	<b>Conclusioni</b>	<b>33</b>



# Capitolo 1

## Nozioni preliminari

### 1.1 Notazione

Siano  $f, g : \mathbb{N} \rightarrow \mathbb{R}$ . Per  $n \rightarrow +\infty$ :

$$f = O(g) \iff \exists c > 0, \exists \bar{n} \in \mathbb{N} \mid \forall n \geq \bar{n}, f(n) \leq cg(n)$$

$$f = o(g) \iff \frac{f(n)}{g(n)} \rightarrow 0$$

$$f = \omega(g) \iff \frac{f(n)}{g(n)} \rightarrow +\infty \iff g = o(f)$$

$$f \sim g \iff 0 < \liminf \left| \frac{f(n)}{g(n)} \right| \leq \limsup \left| \frac{f(n)}{g(n)} \right| < +\infty$$

### 1.2 Grafi

**Definizione 1.1** (Grafo). Un grafo è una coppia ordinata  $(N, G)$  dove  $N = \{1, 2, \dots, n\}$  e  $G = (g_{ij})$  è una matrice quadrata binaria di ordine  $n$  che rappresenta una determinata relazione su  $N$ :

$$g_{ij} = \begin{cases} 1 & \text{se } i \text{ è in relazione con } j \\ 0 & \text{altrimenti} \end{cases}$$

Gli elementi di  $N$  sono detti nodi.

Se il nodo  $i$  è in relazione con il nodo  $j$  si dice che vi è un arco da  $i$  a  $j$ .

Infine,  $G$  è detta matrice delle adiacenze.

**Osservazione 1.1.** Un altro modo per definire un grafo è tramite una coppia  $(N, S)$ , dove  $N$  è un insieme come sopra e  $S = \{(i, j) \mid i, j \in N, i \text{ è in relazione con } j\} \subseteq N \times N$ .

**Definizione 1.2** (Grafo non diretto). Quando la matrice delle adiacenze  $G$  è simmetrica, il grafo si dice non diretto o non orientato.

In questo caso, due nodi in relazione tra loro sono detti adiacenti.

**Definizione 1.3** (Sottografo). Un sottografo di  $(N, G)$  è un grafo  $(\tilde{N}, \tilde{G})$  tale che  $\tilde{N} = \{x_1, \dots, x_k\} \subseteq N$  e  $\tilde{G} = (\tilde{g}_{x_i x_j})$  è una matrice di ordine  $k$  per cui  $\tilde{g}_{ij} = g_{x_i x_j}$ .

**Osservazione 1.2.** D'ora in avanti, si considereranno i grafi non diretti  $(N, G)$  in cui  $N \neq \emptyset$  e  $G \neq (0)$ .

### 1.3 Connessione

**Definizione 1.4** (Vicinato e grado). Dati un grafo  $(N, G)$  e un nodo  $i \in N$ , il vicinato di  $i$  è l'insieme dei nodi adiacenti ad esso, ovvero l'insieme  $N_i(G) = \{j \in N \mid g_{ij} = 1\}$ . Per convenzione,  $i \notin N_i(G)$ .

Il grado del nodo  $i$  è

$$d_i = |N_i(G)| = \sum_{j \neq i} g_{ij}$$

$$0 \leq d_i \leq n - 1$$

**Definizione 1.5** (Percorso). Dati due nodi  $i, j \in N$ , un percorso di estremi  $i$  e  $j$  è un vettore  $w_{ij} = (x_0, x_1, \dots, x_k)$  tale che  $x_0 = i$ ,  $x_k = j$ ,  $x_h \in N$  e  $g_{x_h x_{h+1}} = 1$  per  $h = 0, 1, \dots, k-1$ .

In altri termini, è una sequenza di nodi nella quale ogni elemento è adiacente al precedente e al successivo.

**Definizione 1.6** (Ciclo). Un percorso in cui  $i = j$ , ovvero in cui i due estremi coincidono, si dice ciclo. Il ciclo è semplice se  $i = j$  è l'unico nodo che si ripete all'interno della sequenza.

**Definizione 1.7** (Cammino). Un cammino è un percorso  $w_{ij} = (x_0, x_1, \dots, x_k)$  in cui tutti i nodi sono distinti:  $x_h \neq x_l$  per  $h \neq l$ .

La lunghezza del cammino è  $L(w_{ij}) = |w_{ij}| - 1 = k$ .

**Definizione 1.8** (Geodetica e distanza). Un cammino  $\tilde{w}_{ij}$  di lunghezza minima tra  $i$  e  $j$  si dice geodetica.

La distanza tra i due nodi è la lunghezza di una tale geodetica  $D(i, j) = L(\tilde{w}_{ij}) = \min\{L(w_{ij}) \mid w_{ij} \text{ cammino}\}$ .

**Definizione 1.9** (Connessione). Due nodi  $i$  e  $j$  si dicono connessi se esiste un percorso  $w_{ij}$  di estremi  $i$  e  $j$ .

Analogamente, un grafo è connesso se tutti i suoi elementi sono connessi tra loro, cioè se  $\forall i, j \in N, \exists w_{ij}$  percorso tra  $i$  e  $j$ .

**Definizione 1.10** (Componente connessa). Una componente connessa (o, semplicemente, componente) di un grafo è un sottografo  $(C, G_c)$  tale che

$$\begin{cases} i, j \in C \implies i \text{ e } j \text{ connessi} \\ i \in C, k \notin C \implies i \text{ e } k \text{ non connessi} \end{cases}$$

Se il grafo è connesso, esiste una sola componente che coincide con il grafo stesso.

**Proposizione 1.1.** Sia  $G^m = (g_{ij}^m)$ , con  $1 \leq m \leq n$ . Allora  $g_{ij}^m$  è il numero di percorsi di lunghezza  $m$  tra  $i$  e  $j$ :

$$g_{ij}^m = \text{card}\{w_{ij} \text{ percorso} \mid L(w_{ij}) = m\}$$

*Dimostrazione.* Per induzione su  $m$ :

Per  $m = 1$ ,  $G^1 = G = (g_{ij})$  con  $g_{ij} = 1$  se vi è un arco tra  $i$  e  $j$  (unico percorso di lunghezza 1 tra i due nodi),  $g_{ij} = 0$  altrimenti.

Si supponga che la proposizione sia valida per  $G^{m-1}$ . Allora  $G^m = GG^{m-1}$  e

$$g_{ij}^m = \sum_{k=1}^n g_{ik}g_{kj}^{m-1}$$

Fissato  $k$ , per ipotesi induttiva  $g_{kj}^{m-1} = n_k =$  numero di percorsi  $(k, \dots, j)$  di lunghezza  $m - 1$ .

$$g_{ik}g_{kj}^{m-1} = g_{ik}n_k = \tilde{n}_k = \begin{cases} n_k & \text{se esiste un arco tra } i \text{ e } k \\ 0 & \text{altrimenti} \end{cases}$$

Dunque, per  $k = 1, \dots, n$ ,  $\tilde{n}_k =$  numero di percorsi  $(i, k, \dots, j)$  di lunghezza  $m$ .

$$g_{ij}^m = \sum_{k=1}^n \tilde{n}_k = \text{numero di percorsi } (i, \dots, j) \text{ di lunghezza } m$$

## 1.4 Clustering

Dato un nodo  $i \in N$ , sia  $F_i = \{(k, j) \mid k, j \in N_i(G), g_{kj} = 1\}$  l'insieme delle coppie di nodi del vicinato di  $i$  tra cui vi è un arco.

**Definizione 1.11** (Clustering). Per  $i \in N$  tale che  $|N_i(G)| \geq 2$ , si definisce clustering di  $i$  la quantità

$$Cl_i(G) = \frac{|F_i|}{\text{card}\{(k, j) \mid k, j \in N_i(G)\}}$$

Il concetto di clustering dà una misura della densità di un grafo a livello locale, in quanto rappresenta il rapporto tra il numero delle coppie di nodi adiacenti a  $i$  che sono anche adiacenti tra loro e il numero di tutte le coppie di nodi adiacenti a  $i$ . In altri termini,  $Cl_i(G)$  dà la frequenza di archi tra i nodi adiacenti a  $i$ .

**Definizione 1.12** (Clustering medio). Si definisce clustering medio di un grafo la media dei clustering dei singoli nodi.

$$Cl_{avg}(G) = \frac{1}{n} \sum_{i=1}^n Cl_i(G)$$

**Definizione 1.13** (Clustering globale). Il clustering globale di un grafo è

$$Cl(G) = \frac{\sum_i |F_i|}{\sum_i \text{card}\{(k, j) \mid k, j \in N_i(G)\}}$$

**Osservazione 1.3.** In generale,  $Cl_{avg}(G) \neq Cl(G)$ .

**Osservazione 1.4.** Poiché  $\text{card}\{(k, j) \mid k, j \in N_i(G)\} = \frac{1}{2} d_i(d_i - 1)$ , clustering, clustering medio e clustering globale possono essere espressi come

$$\begin{aligned} Cl_i(G) &= \frac{2|F_i|}{d_i(d_i - 1)} \\ Cl_{avg}(G) &= \frac{2}{n} \sum_{i=1}^n \frac{|F_i|}{d_i(d_i - 1)} \\ Cl(G) &= \frac{2 \sum_i |F_i|}{\sum_i d_i(d_i - 1)} \end{aligned}$$

## 1.5 Proprietà

Dato  $N = \{1, 2, \dots, n\}$ , si denoti l'insieme di tutti i possibili grafi non diretti su  $N$  con  $G(N) = \{(N, G) \text{ grafo non diretto}\}$ .

**Definizione 1.14** (Proprietà). Si dice proprietà di  $N$  un sottoinsieme di grafi  $X \subseteq G(N)$ .

**Definizione 1.15** (Proprietà monotona). Siano  $(N, G), (N, \tilde{G}) \in G(N)$  tali che  $g_{ij} = 1 \implies \tilde{g}_{ij} = 1$ , cioè tali che in  $(N, \tilde{G})$  compaiono tutti gli archi presenti in  $(N, G)$ .

Una proprietà  $X$  è monotona se  $(N, G) \in X \implies (N, \tilde{G}) \in X$ .

Dunque se un grafo  $(N, G)$  appartiene ad una proprietà monotona  $X$ , allora tutti i grafi in cui compaiono gli archi di  $(N, G)$  appartengono alla proprietà.

**Osservazione 1.5.** Sono proprietà monotone:

- esistenza di un arco:  $X_1 = \{(N, G) \mid \exists i, j, i \neq j, g_{ij} = 1\}$
- esistenza di 3 nodi connessi:  $X_2 = \{(N, G) \mid \exists i, j, k, i \neq j \neq k, g_{ij} = g_{jk} = 1\}$
- esistenza di un ciclo:  $X_3 = \{(N, G) \mid \exists w_{ii} = (i, \dots, i) \text{ ciclo}\}$
- connessione:  $X_4 = \{(N, G) \text{ connesso}\}$

# Capitolo 2

## Il modello ER

### 2.1 Grafo di Erdős-Rényi

Il modello di Erdős-Rényi (o, in breve, modello ER) è uno dei modelli base per i grafi aleatori. Si tratta di un grafo con  $n$  nodi in cui ogni arco viene formato indipendentemente con una probabilità  $p$  fissata.

**Definizione 2.1** (Grafo ER). Un grafo ER, denotato con  $G(n, p)$ , è un grafo  $(N, G)$  con  $N = \{1, 2, \dots, n\}$  e in cui la matrice delle adiacenze  $G = (g_{ij})$  è tale che, per ogni coppia di nodi distinti  $(i, j)$ , la probabilità che  $g_{ij} = 1$  è  $P(g_{ij} = 1) = p$ , con  $p \in (0, 1)$  fissato. La diagonale principale è nulla.

Dunque

$$g_{ii} = 0$$
$$g_{ij} = \begin{cases} 1 & \text{con probabilità } p \\ 0 & \text{con probabilità } 1 - p \end{cases}$$

Variando la scelta dei valori di  $n$  e  $p$ , il grafo  $G(n, p)$  assume configurazioni differenti.

Generalmente si considerano grafi ER in cui  $n \rightarrow +\infty$  e  $p = p(n) \rightarrow 0$ .

## 2.2 Grado dei nodi

### 2.2.1 Caso generale

**Teorema 2.1.** Si considerino un grafo  $G(n, p)$  e un suo nodo  $i$ .

Per  $k = 0, \dots, n-1$ , la probabilità che  $i$  abbia grado  $d_i = k$  è binomiale.

Posto  $P_k = P(d_i = k)$ , risulta:

$$P_k = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

*Dimostrazione.* Fissati il nodo  $i$  e  $k$ , con  $0 \leq k \leq n-1$ ,

$$d_i = \sum_{j \neq i} g_{ij} = k$$

se e solo se  $k$  addendi hanno valore 1 e i restanti  $n-1-k$  hanno valore 0, cioè se e solo se

$$\begin{cases} g_{ih_1} = \dots = g_{ih_k} = 1 \\ g_{ih_{k+1}} = \dots = g_{ih_{n-1-k}} = 0 \end{cases}$$

Essendo le probabilità indipendenti, si ha:

$$\begin{aligned} P((g_{ih_1}, \dots, g_{ih_k}, g_{ih_{k+1}}, \dots, g_{ih_{n-1}}) = (1, \dots, 1, 0, \dots, 0)) &= \\ &= \left( \prod_{t=1}^k P(g_{ih_t} = 1) \right) \left( \prod_{t=k+1}^{n-1} P(g_{ih_t} = 0) \right) = \left( \prod_{t=1}^k p \right) \left( \prod_{t=k+1}^{n-1} (1-p) \right) = \\ &= p^k (1-p)^{n-1-k} \end{aligned}$$

Infine, le possibili combinazioni di  $k$  elementi  $(g_{ih_1}, \dots, g_{ih_k})$  scelti tra  $n-1$  sono



$$\binom{n-1}{k} = \frac{(n-1)!}{k!(n-1-k)!}$$

*Dunque*

$$\begin{aligned} P_k &= \binom{n-1}{k} P((g_{ih_1}, \dots, g_{ih_k}, g_{ih_{k+1}}, \dots, g_{ih_{n-1}}) = (1, \dots, 1, 0, \dots, 0)) = \\ &= \binom{n-1}{k} p^k (1-p)^{n-1-k} \end{aligned}$$

**Corollario 2.1.** Il grado  $d_i$  è una variabile aleatoria discreta che assume valori tra 0 e  $n-1$ . Il valore atteso è  $E[d_i] = (n-1)p$ , mentre la varianza è  $V[d_i] = 1-p$ .

*Dimostrazione.* Il valore atteso è dato da:

$$E[d_i] = \sum_{k=1}^{n-1} k P_k = \sum_{k=1}^{n-1} k \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

con

$$k \binom{n-1}{k} = \frac{k(n-1)!}{k!(n-1-k)!} = \frac{(n-1)(n-2)!}{(k-1)!((n-2)-(k-1))!} = (n-1) \binom{n-2}{k-1}$$

*Dunque*

$$\begin{aligned}
E[d_i] &= \sum_{k=1}^{n-1} (n-1) \binom{n-2}{k-1} p^k (1-p)^{n-1-k} = \\
&= (n-1)p \sum_{k=1}^{n-1} \binom{n-2}{k-1} p^{k-1} (1-p)^{(n-2)-(k-1)} = \\
&= (n-1)p \sum_{j=0}^{n-2} \binom{n-2}{j} p^j (1-p)^{n-2-j} = \\
&= (n-1)p(p + (1-p))^{n-2} = (n-1)p
\end{aligned}$$

Analogamente, si dimostra che la varianza è

$$V[d_i] = E[d_i^2] - (E[d_i])^2 = (n^2 p^2 + np(1-p)) - (np)^2 = np(1-p)$$

### 2.2.2 Grafo con $n$ grande e $p$ piccolo

Si consideri ora un grafo  $G(n, p)$  in cui il numero di nodi  $n$  è molto grande e la probabilità  $p$  molto piccola. In questo caso, la distribuzione del grado  $d_i$  è approssimabile a una distribuzione di Poisson con media  $\lambda = np$ .

**Teorema 2.2.** Sia  $G(n, p)$  tale che  $p = p(n) = O(\frac{1}{n})$ .

Sia poi  $P_k = P(d_i = k)$  per un nodo  $i$  fissato. Posto  $\lambda = np$ , risulta:

$$P_k \longrightarrow \frac{\lambda^k}{k!} e^{-\lambda}$$

per  $n \rightarrow +\infty$ .

*Dimostrazione.*

$$\begin{aligned}
P_k &= \binom{n-1}{k} p^k (1-p)^{n-1-k} = \frac{(n-1)!}{k!(n-1-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-1-k} = \\
&= \frac{\lambda^k}{k!} \cdot \frac{(n-1)!}{n^k(n-1-k)!} \left( \left(1 - \frac{\lambda}{n}\right)^{-\frac{n}{\lambda}} \right)^{-\lambda} \left(1 - \frac{\lambda}{n}\right)^{-(k+1)}
\end{aligned}$$

Per  $n \rightarrow +\infty$  si ha:

$$\begin{aligned}\frac{(n-1)!}{n^k(n-1-k)!} &\longrightarrow 1 \\ \left(1 - \frac{\lambda}{n}\right)^{-\frac{n}{\lambda}} &\longrightarrow e \\ 1 - \frac{\lambda}{n} &\longrightarrow 1\end{aligned}$$

Allora

$$P_k \longrightarrow \frac{\lambda^k}{k!} \cdot 1 \cdot e^{-\lambda} \cdot 1^{-(k+1)} = \frac{\lambda^k}{k!} e^{-\lambda}$$

**Corollario 2.2.** Nelle ipotesi del teorema, il valore atteso e la varianza del grado  $d_i$  sono rispettivamente  $E[d_i] = \lambda$  e  $V[d_i] = \lambda$ .

*Dimostrazione.* Valore atteso:

$$E[d_i] = \sum_{k=1}^{+\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

Varianza:

$$\begin{aligned}
E[d_i^2] &= \sum_{k=1}^{+\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} = \sum_{k=1}^{+\infty} (k(k-1) + k) e^{-\lambda} \frac{\lambda^k}{k!} = \\
&= \sum_{k=1}^{+\infty} k(k-1) e^{-\lambda} \frac{\lambda^k}{k!} + \sum_{k=1}^{+\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \\
&= \lambda^2 e^{-\lambda} \sum_{k=1}^{+\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda = \lambda^2 + \lambda \\
V[d_i] &= E[d_i^2] - (E[d_i])^2 = \lambda^2 + \lambda - \lambda^2 = \lambda
\end{aligned}$$

## 2.3 Clustering

Con considerazioni analoghe a quelle per il grado di un nodo  $i \in N$ , si può dedurre il valore atteso del clustering

$$Cl_i(G) = \frac{2}{d_i(d_i - 1)} \cdot |F_i|$$

dove  $F_i = \{(k, j) \mid k, j \in N_i(G), g_{k,j} = 1\}$

Poiché le possibili coppie  $(k, j)$  di nodi adiacenti a  $i$  sono  $\frac{1}{2} d_i(d_i - 1)$ , segue che

$$\begin{aligned}
0 &\leq |F_i| \leq \frac{d_i(d_i - 1)}{2} \\
0 &\leq Cl_i(G) \leq 1
\end{aligned}$$

**Proposizione 2.1.** Il valore atteso del clustering di un nodo è  $E[Cl_i(G)] = p$ .

*Dimostrazione.* Supponendo che  $d_i$  sia noto, sia  $A = \frac{1}{2} d_i(d_i - 1)$ . Per  $0 \leq h \leq A$ , la probabilità che  $|F_i| = h$  è binomiale.

$$\begin{aligned}
 P(|F_i| = h) &= \binom{A}{h} p^h (1-p)^{A-h} \\
 E[|F_i|] &= Ap \\
 E[Cl_i(G)] &= E\left[\frac{|F_i|}{A}\right] = \frac{1}{A} E[|F_i|] = p
 \end{aligned}$$

**Osservazione 2.1.** Se  $p = p(n) \rightarrow 0$  per  $n \rightarrow +\infty$ , allora anche  $E[Cl_i(G)] = p \rightarrow 0$ .

## 2.4 Componente gigante

**Definizione 2.2** (Componente gigante). Si consideri un grafo  $G(n, p)$  con  $n \rightarrow +\infty$ . Si dice che il grafo ha una componente gigante se esiste una componente connessa  $(C, G_c)$  tale che  $|C| \sim n$ .

Dunque una componente gigante è una componente connessa che contiene una grande percentuale dei nodi; la sua dimensione tende all'infinito.

In genere, se la componente gigante esiste è unica.

**Proposizione 2.2.** La probabilità  $P(i \in C)$  che un nodo appartenga dalla componente gigante è radice di:

$$f(x) = (1 - px)^{n-1} + x - 1$$

Per  $x \in (0, 1)$ , esiste una sola radice  $\bar{x} \in \left(\frac{1}{p}, 1\right)$ . Inoltre  $\bar{x} \rightarrow 1$ .

*Dimostrazione.* Un nodo è al di fuori della componente gigante se e solo se lo sono tutti i nodi ad esso adiacenti. Allora  $P_1 = P_2$  con

$$\begin{aligned}
P_1 &= P(i \notin \text{componente gigante}) = 1 - x \\
P_2 &= P(\forall j \in N_i(G), j \notin \text{componente gigante}) = \\
&= \sum_{k=0}^{n-1} P(d_i = k, \text{ e } h_1, \dots, h_k \text{ adiacenti } \notin \text{componente gigante}) = \\
&= \sum_{k=0}^{n-1} P_k \cdot P(h_1, \dots, h_k \notin \text{componente gigante}) = \\
&= \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k} (1-x)^k = \\
&= \sum_{k=0}^{n-1} \binom{n-1}{k} (p-px)^k (1-p)^{n-1-k} = \\
&= (p-px + 1-p)^{n-1} = (1-px)^{n-1} \\
P_1 = P_2 &\iff 1-x = (1-px)^{n-1} \iff f(x) = (1-px)^{n-1} + x - 1 = 0
\end{aligned}$$

*Inoltre*

$$\begin{cases} f\left(\frac{1}{p}\right) = \frac{1}{p} - 1 < 0 \\ f(1) = (1-p)^{n-1} > 0 \end{cases} \implies \frac{1}{p} < \bar{x} < 1$$

con  $\frac{1}{p} \rightarrow 1$ .

**Osservazione 2.2.** L'esistenza della componente gigante è una proprietà monotona.

## 2.5 Proprietà e funzioni soglia

**Definizione 2.3** (Funzione soglia). Sia  $G(n, p)$  un grafo ER con  $p = p(n)$ .

Siano poi  $X$  una proprietà monotona e  $P(X) = P(G(n, p) \in X)$ .

Una funzione soglia per  $X$  è una funzione  $t : \mathbb{N} \rightarrow \mathbb{R}$  tale che

$$\begin{cases} p = o(t) \implies P(X) \rightarrow 0 \\ p = \omega(t) \implies P(X) \rightarrow 1 \end{cases}$$

per  $n \rightarrow +\infty$ .

Quindi in  $t(n)$  avviene una transizione di fase: la probabilità di  $X$  tende a 0 se  $p$  decresce più velocemente di  $t(n)$ , e tende a 1 se  $p$  decresce più lentamente.

**Proposizione 2.3.** Data la proprietà  $X = \{(N, G) \mid \text{esiste almeno un arco}\} = \{G(n, p) \mid \exists i, j, i \neq j, g_{ij} = 1\}$ , una funzione soglia è

$$t(n) = \frac{1}{n^2}$$

*Dimostrazione.*  $G$  è una matrice simmetrica, dunque le possibili coppie di nodi distinti  $(i, j)$  sono  $\frac{n(n-1)}{2} \sim n^2$ . Ogni coppia ha probabilità di avere un arco  $P(g_{ij} = 1) = p$ , e tali probabilità sono indipendenti.

Inoltre  $P(X) = 1 - P(\bar{X})$ , dove  $\bar{X} = \{G(n, p) \mid \text{non esistono archi}\} = \{G(n, p) \mid g_{ij} = 0 \forall i, j\}$ .

$$P(\bar{X}) = (1 - p)^{\frac{n(n-1)}{2}}$$

$$P(X) = 1 - (1 - p)^{\frac{n(n-1)}{2}} \sim 1 - (1 - p)^{n^2}$$

*Dunque*

$$\begin{cases} pn^2 \rightarrow 0 \implies (1-p)^{n^2} \rightarrow 1 \implies P(X) \rightarrow 1-1=0 \\ pn^2 \rightarrow +\infty \implies (1-p)^{n^2} \rightarrow 0 \implies P(X) \rightarrow 1-0=1 \end{cases}$$

**Osservazione 2.3.** In un grafo  $G(n, p)$  con  $p = p(n) \sim \frac{1}{n^2}$ , il valore atteso del grado di un nodo è

$$E[d_i] = (n-1)p \sim \frac{n}{n^2} = \frac{1}{n}$$

Analogamente, si dimostrano le seguenti proposizioni.

**Proposizione 2.4.** Una funzione soglia per la proprietà  $X = \{G(n, p) \mid \text{esiste una componente connessa con almeno 3 nodi}\}$  è

$$t(n) = \frac{1}{n^{3/2}} = \frac{1}{n\sqrt{n}}$$

Se  $p = p(n) \sim \frac{1}{n\sqrt{n}}$ , il valore atteso del grado  $d_i$  è  $E[d_i] \sim \frac{1}{\sqrt{n}}$ .

**Proposizione 2.5.** Una funzione soglia per  $X = \{G(n, p) \mid \text{esiste almeno un ciclo}\}$  è

$$t(n) = \frac{1}{n}$$

Se  $p = p(n) \sim \frac{1}{n}$ , il valore atteso del grado è  $E[d_i] \sim 1$ .

**Proposizione 2.6.** La funzione  $t(n) = \frac{1}{n}$  è funzione soglia anche per  $X = \{G(n, p) \mid \text{esiste una componente gigante}\}$ .



**Proposizione 2.7.** Una funzione soglia per  $X = \{G(n, p) \text{ connesso}\}$  è

$$t(n) = \frac{\log n}{n}$$

Se  $p = p(n) \sim \frac{\log n}{n}$ , il valore atteso del grado è  $E[d_i] \sim \log n$ .

## 2.6 Configurazioni di un grafo ER

Si consideri un grafo  $G(n, p)$  con  $p = p(n)$  e  $n \rightarrow +\infty$ .

- $p \sim \frac{1}{n^2\sqrt{n}} \implies$  nodi isolati,  $E[d_i] \sim \frac{1}{n\sqrt{n}}$
- $p \sim \frac{1}{n^2} \implies$  soglia per la presenza di un arco,  $E[d_i] \sim \frac{1}{n}$
- $p \sim \frac{1}{n\sqrt{n}} \implies$  archi, soglia per 3 nodi connessi,  $E[d_i] \sim \frac{1}{\sqrt{n}}$
- $p \sim \frac{1}{n} \implies$  archi e gruppi di almeno 3 nodi connessi, soglia per un ciclo e per la componente gigante,  $E[d_i] \sim 1$
- $p \sim \frac{\sqrt{\log n}}{n} \implies$  archi, gruppi di almeno 3 nodi connessi, cicli e componente gigante,  $E[d_i] \sim \sqrt{\log n}$
- $p \sim \frac{\log n}{n} \implies$  soglia per la connessione,  $E[d_i] \sim \log n$
- $p \sim \frac{\log n}{\sqrt{n}} \implies$  il grafo è connesso,  $E[d_i] \sim \sqrt{n} \log n$

Di seguito sono riportati alcuni esempi di grafi ER con  $n = 100$ .

Figura 2.1: All'aumentare di  $p$  compaiono i primi archi.

(a)  $p=0,00001$  - nodi isolati      (b)  $p=0,0001$  - soglia per un arco

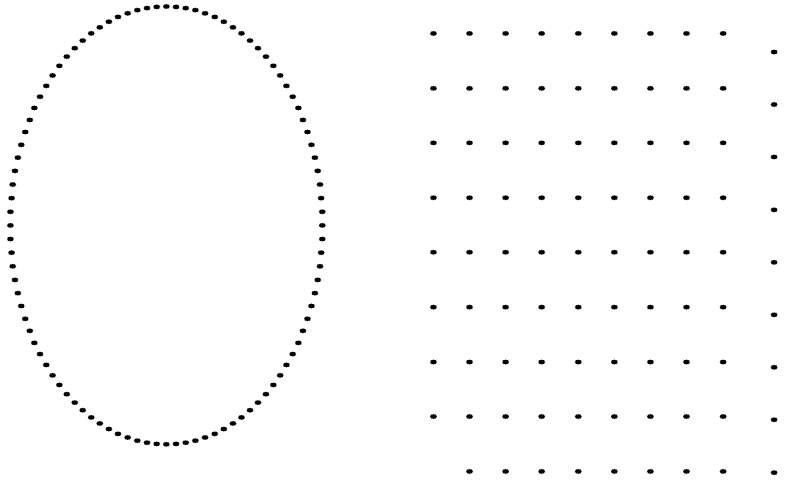


Figura 2.2: Poi gruppi di 3 nodi connessi.

(a)  $p=0,001$  - soglia per 3 nodi connessi      (b)  $p=0,01$  - soglia per un ciclo e la componente gigante

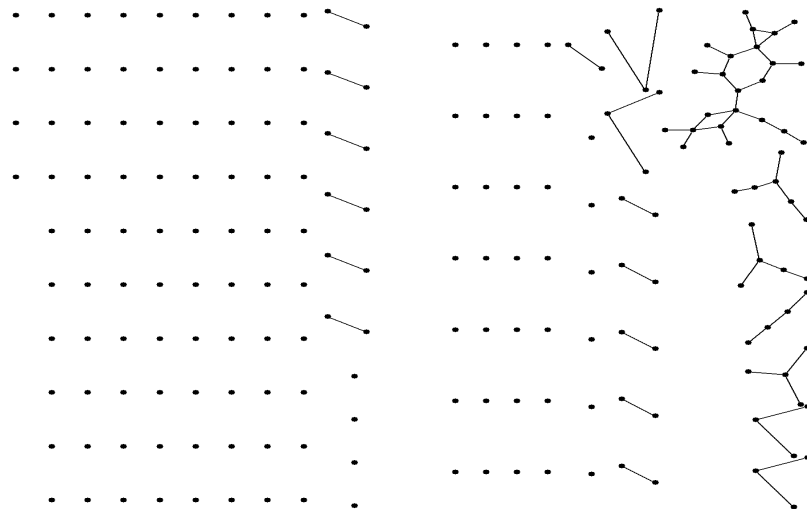
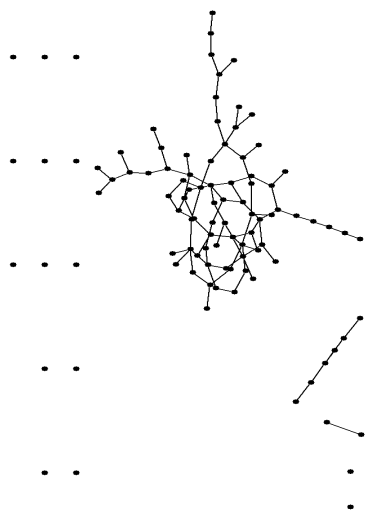


Figura 2.3: Cicli e la componente gigante.

(a)  $p=0,0215$  - cicli e componente  
gigante



(b)  $p=0,0465$  - soglia per la  
connessione

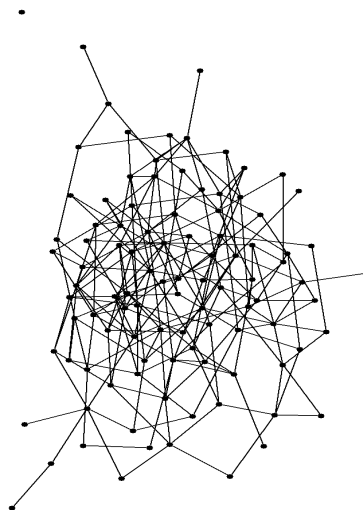
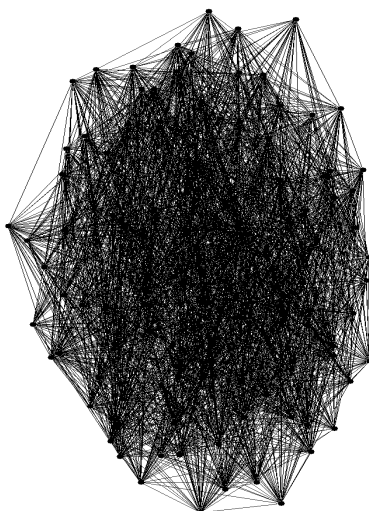


Figura 2.4: Infine, il grafo è connesso.

(a)  $p=0,465$  - connessione





## Capitolo 3

# Esempio di applicazione: diffusione

Una possibile applicazione di quanto visto è la modellizzazione del processo di trasmissione di una determinata informazione. In questo caso, si studia un grafo in cui ogni arco rappresenta il passaggio di tale informazione da un nodo all'altro.

In particolare, si esaminerà la diffusione di una malattia infettiva all'interno di una popolazione numerosa, a partire da uno o più individui infetti.

Si vedrà che nel grafo considerato gli archi rappresentano (in maniera più o meno esplicita) il contagio da un nodo all'altro, mentre ad ogni nodo viene assegnato uno stato: immune, sano o infetto.

Dalla struttura del grafo, è possibile dedurre l'andamento del contagio.

### 3.1 Diffusione su grafo ER

Si consideri un grafo ER che modella le interazioni di un insieme molto grande di persone, ovvero un grafo  $G_1 = G(n, p)$  in cui ogni nodo e ogni arco rappresentano rispettivamente un individuo e il contatto tra due individui.

Si consideri ora una malattia che si trasmette con una determinata probabilità  $f \in (0, 1]$  quando due individui vengono in contatto.

Sia  $\pi \in [0, 1)$  la percentuale di persone naturalmente immuni al contagio: per ogni nodo,  $\pi$  rappresenta la probabilità che questo sia immune.

Si supponga infine che vi sia un singolo individuo già infetto, scelto in maniera casuale tra quelli non immuni.

Per modellizzare il contagio sulla popolazione rappresentata da  $G_1$ , si costruisce un grafo  $G_2$  modificando alcuni aspetti di  $G_1$  stesso:

- si elimina una percentuale  $\pi$  dei nodi (indipendentemente dagli altri, ogni nodo ha probabilità  $\pi$  di essere eliminato)
- si elimina una percentuale  $1 - f$  degli archi
- uno degli  $n(1 - \pi)$  nodi rimasti assume lo stato infetto, mentre gli altri sono sani

Si ottiene così un grafo  $G_2 = G(n(1 - \pi), pf)$  in cui un singolo nodo è infetto.

Nel grafo compaiono soltanto i nodi non immuni, mentre gli archi rappresentano la possibilità di contagio: se un nodo è infetto, i nodi ad esso adiacenti diventano infetti.

Dunque tutti i nodi connessi a quello inizialmente infetto, cioè tutti quelli appartenenti alla sua stessa componente, vengono contagiati.

Poiché  $\pi$  e  $f$  sono costanti indipendenti da  $n$  con  $(1 - \pi), f \in (0, 1]$ , le caratteristiche asintotiche (per  $n \rightarrow +\infty$ ) del grafo restano invariate.

In particolare, se

$$p \in \left( \frac{1}{n}, \frac{\log n}{n} \right)$$

si ottiene un grafo non connesso in cui è presente la componente gigante.

In questo caso, se la componente gigante viene infettata si parla di contagio non banale; in caso contrario, il contagio può essere considerato trascurabile. La probabilità di un contagio non banale coincide con la probabilità  $\bar{x}$  che il nodo da cui parte tale contagio appartenga alla componente gigante.

Come già visto, risulta:

$$\frac{1}{pf} < \bar{x} < 1$$
$$\bar{x} \longrightarrow 1 \text{ per } n \rightarrow +\infty$$

## 3.2 Modello SIS

Si vuole modellizzare la diffusione in dipendenza dal tempo, ovvero una situazione in cui lo stato di ogni individuo (sano o infetto) e le relazioni che si stabiliscono all'interno della popolazione cambiano nel tempo.

Si suppone che la probabilità che una persona venga contagiata sia proporzionale al numero di persone infette con cui questa viene in contatto.

Inoltre, si tiene conto di un tasso di recessione spontanea.

**Definizione 3.1** (Approssimazione di campo medio). Si considerino un modello dipendente dal tempo e una variabile  $\alpha = \alpha(t)$ . Si dice approssimazione di campo medio la tecnica di dedurre la variazione  $\Delta\alpha$  tramite approssimazione continua nel tempo.

Se, ad ogni tempo  $t$  fissato, il valore atteso di  $\Delta\alpha$  è dato da un'espressione  $f(t)$ , un'approssimazione continua in  $t$  permette di scrivere:

$$\frac{d\alpha(t)}{dt} = f(t)$$

**Definizione 3.2** (Fase stazionaria). In un modello variabile nel tempo, si ha una fase stazionaria rispetto a una variabile  $\alpha = \alpha(t)$  quando il suo valore atteso rimane costante, ovvero quando

$$\frac{d\alpha(t)}{dt} = 0$$

**Definizione 3.3** (Modello SIS). Il modello SIS (Susceptible Infected Susceptible) è il dato di un insieme di nodi  $N = \{1, 2, \dots, n\}$ , ciascuno dei quali è nello stato sano oppure infetto.

Ad ogni tempo  $t$ , su  $N$  è definito un grafo aleatorio  $G_t$ .

Si considera un'infezione per la quale:

- inizialmente vi è una percentuale  $\rho(0)$  di nodi infetti
- la probabilità che un nodo sano sia contagiato è proporzionale al numero di adiacenti infetti con un tasso  $v > 0$
- la probabilità che un nodo infetto guarisca è  $\delta > 0$

Si vuole modellizzare il decorso dell'infezione, e studiare le condizioni di esistenza per una fase stazionaria.

### 3.2.1 Costruzione del modello

Poiché la diffusione ha un legame molto stretto con il grado dei nodi, è utile considerare la probabilità che un nodo di grado  $k$  sia infetto

$$\rho_k = \rho_k(t) = P(i \text{ infetto al tempo } t \mid d_i(t) = k)$$

In altri termini,  $\rho_k \in [0, 1]$  è la percentuale di nodi infetti tra quelli di grado



$k$ , con  $k = 0, 1, \dots, n - 1$ .

Sia dunque  $\rho_k(0)$  la percentuale di nodi di grado  $k$  da cui parte il contagio.

Si definisce poi  $\theta = \theta(t)$  la probabilità di incontrare un nodo infetto: per ogni nodo  $i \in N$ ,  $\theta \in [0, 1]$  rappresenta la probabilità che nel vicinato di  $i$  sia presente un nodo infetto. Risulta:

$$\theta = \sum_{h=1}^{n-1} \frac{hP_h}{E[d_i]} \rho_h$$

dove  $P_h$  è la probabilità che un nodo abbia grado  $h$ .

**Dimostrazione.** La probabilità che  $i$  incontri un nodo infetto è

$$\begin{aligned} \theta &= \sum_{h=0}^{n-1} P(i \text{ incontra un nodo di grado } h \text{ infetto}) = \\ &= \sum_{h=0}^{n-1} P(i \text{ incontra un nodo di grado } h) \cdot P(\text{il nodo di grado } h \text{ è infetto}) = \\ &= \sum_{h=0}^{n-1} \frac{hP_h}{E[d_i]} \cdot \rho_h = \sum_{h=1}^{n-1} \frac{hP_h}{E[d_i]} \cdot \rho_h \end{aligned}$$

Ad un tempo  $t$  fissato, si consideri un nodo  $i$  con grado  $d_i(t) = k$ , infetto con probabilità  $\rho_k$ . Si ha:

- la probabilità che  $i$  sia sano e venga infettato è  $(1 - \rho_k)v\theta k$
- la probabilità che  $i$  sia infetto e guarisca è  $\rho_k\delta$

Segue che la variazione nella percentuale  $\rho_k$  di nodi di grado  $k$  infetti è data dalla differenza tra quelli sani che sono stati infettati e quelli infetti che sono guariti.

Dunque per  $k = 0, 1, \dots, n-1$  si ha

$$\begin{aligned} \frac{d\rho_k}{dt} &= (1 - \rho_k)v\theta k - \rho_k\delta = \\ &= vk(1 - \rho_k) \sum_{h=1}^{n-1} \frac{hP_h}{E[d_i]} \rho_h - \rho_k\delta \end{aligned}$$

Si noti che non è necessario ricavare un'espressione esplicita per  $\rho_k$  per determinare le condizioni di esistenza di una fase stazionaria e dare una prima valutazione dell'andamento generale dell'infezione.

### 3.2.2 Fase stazionaria

Un fase stazionaria per la diffusione dell'infezione si ha quando la percentuale di nodi infetti è costante, cioè quando

$$\frac{d\rho_k}{dt} = (1 - \rho_k)v\theta k - \rho_k\delta = 0$$

Questo avviene per

$$\begin{aligned} \rho_k &= \frac{\lambda\theta k}{\lambda\theta k + 1} \\ \text{dove } \lambda &= \frac{v}{\delta} > 0 \end{aligned}$$

In questo caso, risulta

$$\theta = H(\theta) = \sum_{h=1}^{n-1} \frac{h^2 P_h}{E[d_i]} \cdot \frac{\lambda\theta}{\lambda\theta h + 1}$$

con  $H(\theta) : [0, 1] \rightarrow \mathbb{R}$  funzione continua strettamente crescente e concava.

*Dimostrazione.*

$$\begin{aligned}\frac{d\rho_k}{dt} &= (1 - \rho_k)v\theta k - \rho_k\delta = v\theta k - \rho_k(v\theta k + \delta) = 0 \\ \Leftrightarrow \rho_k &= \frac{v\theta k}{v\theta k + \delta} = \frac{\lambda\theta k}{\lambda\theta k + 1}\end{aligned}$$

In questo caso si può esprimere  $\theta$  come

$$\begin{aligned}\theta &= \sum_{h=1}^{n-1} \frac{hP_h}{E[d_i]} \cdot \rho_h = \sum_{h=1}^{n-1} \frac{hP_h}{E[d_i]} \cdot \frac{\lambda\theta h}{\lambda\theta h + 1} = \\ &= \sum_{h=1}^{n-1} \frac{h^2 P_h}{E[d_i]} \cdot \frac{\lambda\theta}{\lambda\theta h + 1} = H(\theta)\end{aligned}$$

$H(\theta)$  è strettamente crescente poiché la derivata prima è positiva:

$$\begin{aligned}H'(\theta) &= \sum_{h=1}^{n-1} \frac{h^2 P_h \lambda}{E[d_i]} \cdot \left( \frac{1}{\lambda\theta h + 1} - \frac{\lambda\theta h}{(\lambda\theta h + 1)^2} \right) = \\ &= \sum_{h=1}^{n-1} \frac{h^2 P_h \lambda}{E[d_i]} \cdot \frac{\lambda\theta h + 1 - \lambda\theta h}{(\lambda\theta h + 1)^2} = \\ &= \sum_{h=1}^{n-1} \frac{h^2 P_h \lambda}{E[d_i](\lambda\theta h + 1)^2} > 0\end{aligned}$$

Infine,  $H(\theta)$  è concava in quanto la derivata seconda è negativa:

$$\begin{aligned}H''(\theta) &= \sum_{h=1}^{n-1} \frac{h^2 P_h \lambda}{E[d_i]} \cdot \frac{-2\lambda h}{(\lambda\theta h + 1)^3} = \\ &= -2 \sum_{h=1}^{n-1} \frac{h^3 P_h \lambda^2}{E[d_i](\lambda\theta h + 1)^3} < 0\end{aligned}$$

In conclusione, si ha una fase stazionaria se  $\theta \in [0, 1]$  è punto fisso di  $H(\theta)$ .

Si osservi che 0 è punto fisso, in quanto  $H(0) = 0$ .

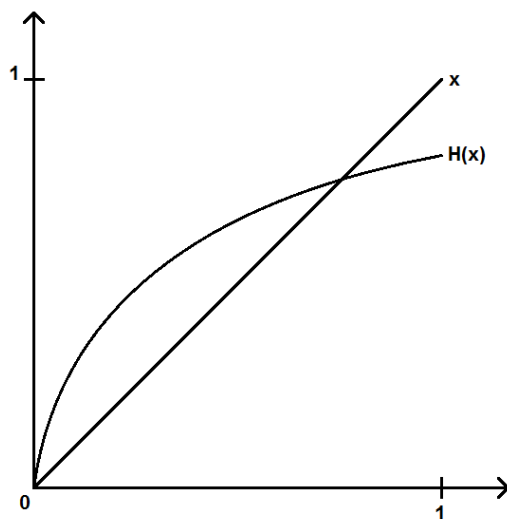
Per  $\theta = 0$ , si ha  $\rho_k(t) = 0$  per ogni  $k$  e per ogni  $t$ : si tratta del caso limite in cui l'infezione non si diffonde poiché tutti i nodi sono sani all'istante iniziale.

Essendo  $H(\theta)$  continua, concava e strettamente crescente su  $[0, 1]$ , esiste un altro punto fisso (non nullo) se e solo se

$$\begin{cases} H'(0) > 1 \\ H(1) \leq 1 \end{cases}$$

cioè se e solo se

$$\frac{v}{\delta} = \lambda > \frac{E[d_i]}{E[d_i^2]}$$



*Dimostrazione.*

$$H'(0) = \sum_{h=1}^{n-1} \frac{h^2 P_h \lambda}{E[d_i]} = \lambda \frac{E[d_i^2]}{E[d_i]} > 1 \iff \lambda > \frac{E[d_i]}{E[d_i^2]}$$

$$H(1) = \sum_{h=1}^{n-1} \frac{h^2 P_h \lambda}{E[d_i](\lambda h + 1)} \leq 1 \text{ sempre vero per } \lambda > 0$$

In definitiva, se il tasso di diffusione è abbastanza grande rispetto a quello di guarigione, cioè se

$$\frac{v}{\delta} > \frac{E[d_i]}{E[d_i^2]}$$

esiste  $\bar{\theta} \in (0, 1)$  tale che  $H(\bar{\theta}) = \bar{\theta}$ . Allora si ha una fase stazionaria in cui

$$\rho_k(t) = \frac{\lambda \bar{\theta} k}{\lambda \bar{\theta} k + 1} = \rho_k(0) \text{ costante}$$

Si ricava, dunque, che il punto fisso  $\bar{\theta}$  è

$$\bar{\theta} = \frac{\rho_k(0)}{\lambda k(1 - \rho_k(0))}$$

Al contrario, se

$$\frac{v}{\delta} \leq \frac{E[d_i]}{E[d_i^2]}$$

risulta  $\theta > H(\theta)$  per  $\theta > 0$ . In questo caso

$$\begin{aligned}\rho_k &> \frac{\lambda\theta k}{\lambda\theta k + 1} \\ \implies \frac{d\rho_k}{dt} &= (1 - \rho_k)v\theta k - \rho_k\delta < 0\end{aligned}$$

Dunque la percentuale di nodi infetti decresce nel tempo.

### 3.2.3 Modello SIS su grafo ER

Si consideri il modello SIS nel caso in cui i grafi aleatori definiti su  $N$  ad ogni istante  $t$  siano un grafi ER:  $G_t = G(n, p)$  con  $p = p(n)$  indipendente da  $t$ . In questo caso si ha:

$$\begin{aligned}P_k &= \binom{n-1}{k} p^k (1-p)^{n-1-k} \\ E[d_i] &= (n-1)p \\ E[d_i^2] &= n^2 p^2\end{aligned}$$

Fissati il tasso di infezione  $v$ , quello di guarigione  $\delta$  e la percentuale di nodi inizialmente infetti  $\rho_k(0)$  per  $k = 0, 1, \dots, n-1$ , la probabilità di incontrare un nodo infetto è

$$\begin{aligned}\theta(t) &= \sum_{h=1}^{n-1} \frac{h}{(n-1)p} \cdot \binom{n-1}{h} p^h (1-p)^{n-1-h} \rho_h(t) = \\ &= \sum_{h=1}^{n-1} \binom{n-2}{h-1} p^{h-1} (1-p)^{n-h-1} \rho_h(t) = \sum_{h=1}^{n-1} P'_{h-1} \rho_h(t)\end{aligned}$$

dove  $P'_j$  è la probabilità che in un grafo  $G' = G(n-1, p)$  un nodo abbia grado  $j$ .

Inoltre

$$\frac{E[d_i]}{E[d_i^2]} = \frac{(n-1)}{n^2 p} \sim \frac{1}{np}$$

Per  $n \rightarrow +\infty$ , risulta:

- se  $\delta < npv$ , il numero di nodi infetti può essere decrescente, stabile (fase stazionaria) o crescente
- se  $\delta \geq npv$ , il numero di nodi infetti è decrescente, e l'unica fase stazionaria si ha per  $\rho(0) = 0$





# Capitolo 4

## Conclusioni

Si è data una definizione formale per il modello ER, e sono stati considerati i grafi  $G(n, p)$  con  $n \rightarrow +\infty$  e  $p = p(n) \rightarrow 0$ .

Si è visto che il grado dei nodi (misura della connessione) risulta essere una variabile aleatoria con distribuzione binomiale, mentre il clustering (misura della densità di archi a livello locale) tende a zero.

Successivamente sono state determinate le funzioni soglia per alcune proprietà monotone particolarmente significative, consentendo così di descrivere diverse configurazioni possibili per un grafo ER al variare di  $p$ .

Infine, si è mostrato come si possano utilizzare i grafi ER per modellizzare la diffusione di una malattia all'interno di una popolazione numerosa.

Sono stati presentati un primo modello di rappresentazione, che sfrutta i grafi ER in maniera particolarmente esplicita, e il modello SIS, che permette di descrivere l'andamento della diffusione mettendone in evidenza la dipendenza dal tempo.



# Bibliografia

- [1] M. O. Jackson, *Social and Economic Networks*, Princeton University Press, Princeton 2010.
- [2] P. Baldi, *Calcolo delle probabilità e statistica*, McGraw-Hill, Milano 1998.
- [3] <http://www.disi.unige.it/person/DellamicoM/teaching/ar2-0607/3-modelli.pdf>