Introduzione al corso di Analisi di Reti

1 Premessa

Quando vengono resi disponibili agli studenti i lucidi delle lezioni, spesso si tende a studiare solo su questi materiali, il che porta ad una preparazione superficiale e insufficiente. Pertanto, il libro di testo e le dispense online sono un necessario complemento ai lucidi per una preparazione adeguata. Nei lucidi sarà indicata la porzione di testo o la dispensa cui far riferimento per ogni argomento, con qualche eccezione: ad esempio, il modello di Erdős-Rényi viene trattato solo nei lucidi.

2 Obiettivo del corso

L'obiettivo del corso è analizzare reti nell'accezione più ampia possibile, considerando prestazioni, struttura, utilizzo, e tecniche trasversali. Il corso è una sintesi tra:

- Matematica e informatica (modelli, analisi, algoritmi, complessità)
- Economia (relazioni come incentivo/disincentivo a comportamenti)
- Scienze sociali (studio di strutture e interazioni in gruppi e popolazioni)

3 Analisi di Reti Mod. 1

Il corso non è puramente informatico (non è un corso di algoritmi) ma si occupa di modelli e della loro analisi. Si generano reti per modellare caratteristiche reali e si dimostra che le reti generate mostrano tali caratteristiche. Si modellano formalmente problemi di gestione delle reti e se ne cercano le soluzioni.

4 Definizione di Rete

Una rete è uno schema di interconnessione tra entità. Esistono:

- Reti fisiche
- Reti sociali
- Reti di informazioni

Si studiano comportamenti aggregati influenzati dai legami fra individui (effetti informativi, fenomeni di diffusione, ricerca di percorsi brevi, fenomeno "rich get richer").

5 Struttura di una rete

Si analizzano proprietà globali delle reti grandi (componenti giganti). Si distinguono nodi centrali, periferici, e si studia la distribuzione dei gradi dei nodi e il diametro del grafo.

6 Studio dei Fenomeni sulle Reti

Talvolta si studiano fenomeni a livello di popolazione, altre volte è necessaria la struttura fisica della rete. Per studiare una rete la si modella tramite un grafo:

- Grafi orientati e non orientati
- Percorsi, alberi ricoprenti, componenti connesse, diametro, BFS

7 Modelli generativi di grafi casuali

Le reti reali sono limitate e difficili da ottenere, quindi spesso si generano reti casuali tramite modelli probabilistici. Si studiano quattro modelli: tra questi il modello di Erdős-Rényi.

8 Il modello di Erdős-Rényi — Componenti Giganti

Definiamo il grafo $G_{n,p}$ come il grafo casuale con nodo set $[n] = \{1, 2, \ldots, n\}$ e con probabilità p ogni arco tra coppie distinte di nodi.

Teorema 1. *Se*

$$p > \frac{\ln 64}{n}$$

allora, con alta probabilità, $G_{n,p}$ contiene una componente connessa costituita da almeno metà dei suoi nodi, cioè

$$P\left(X \geq \frac{n}{2}\right) \geq 1 - 2^{-n/8}$$

dove X è la variabile aleatoria che indica la dimensione della più grande componente connessa.

Lemma 2. *Se*

$$X < \frac{n}{2}$$

allora esiste un insieme $A \subseteq [n]$ tale che

$$\frac{n}{4} \leq |A| < \frac{3n}{4}$$

e non esistono archi tra i nodi di A e i nodi di $[n] \setminus A$.

Proof. Siano $C_1, C_2, \ldots, C_k$ le componenti connesse di $G_{n,p}$, ordinate per cardinalità non decrescente:

$$|C_1| \leq |C_2| \leq \ldots \leq |C_k|$$

Poiché $X < \frac{n}{2}$, allora $|C_i| < \frac{n}{2}$ per ogni $1 \leq i \leq k$.

Scegliamo un indice h tale che

$$|C_1| + |C_2| + \cdots + |C_{h-1}| < \frac{n}{4}$$

e

$$|C_1| + |C_2| + \cdots + |C_{h-1}| + |C_h| \geq \frac{n}{4}$$

Ovviamente $h < k$ perché $|C_k| < \frac{n}{2}$.

Poniamo allora

$$A = C_1 \cup C_2 \cup \cdots \cup C_h$$

Chiaramente

$$|A| \geq \frac{n}{4}$$

e

$$|A| < \frac{n}{4} + \frac{n}{2} = \frac{3n}{4}$$

Per costruzione non esistono archi tra A e $[n] \setminus A$, altrimenti le componenti non sarebbero separate. □

Definiamo un insieme A con queste proprietà come *insieme buono.*

Dimostrazione del Teorema. Calcoliamo la probabilità complementare:

$$P\left(X < \frac{n}{2}\right) \leq P(\exists A \subseteq [n] : A \text{ è buono})$$

Applicando lo Union Bound:

$$P\left(\bigcup_{A\subseteq[n]} \text{“A è buono”}\right) \leq \sum_{A} P\left(\text{non ci sono archi tra } A \text{ e } [n] \setminus A\right)$$

Il numero di archi possibili tra A e $[n] \setminus A$ è

$$|A| \cdot (n - |A|)$$

e la probabilità che ciascuno di questi archi non sia presente è $(1 - p)$.

Quindi

$$P(\text{non ci sono archi tra } A \text{ e } [n]\setminus A) = (1-p)^{|A|(n-|A|)} \leq (1-p)^{\frac{n}{4}\cdot\frac{3n}{4}} = (1-p)^{\frac{3n^2}{16}}$$

Poiché $(1 - p) < 1$, il termine è massimo per il più piccolo valore di $|A|(n - |A|)$, ovvero $n/4 \cdot 3n/4$.

Poiché ci sono al massimo 2^n sottoinsiemi A, si ottiene:

$$P\left(X < \frac{n}{2}\right) \leq 2^n(1-p)^{\frac{3n^2}{16}}$$

Usando la disuguaglianza $1 - x \leq e^{-x}$, sostituiamo:

$$\leq 2^n e^{-p\frac{3n^2}{16}} = 2^n e^{-\frac{3}{16}pn^2}$$

Ponendo

$$p = \frac{\ln 64}{n}$$

otteniamo:

$$P\left(X < \frac{n}{2}\right) \leq 2^n e^{-\frac{3}{16}\ln 64\cdot n} = 2^n 64^{-\frac{3}{16}n} = 2^n 2^{-6\cdot\frac{3}{16}n} = 2^n 2^{-\frac{18}{16}\cdot n} = 2^{-n/8}$$

che tende a zero esponenzialmente al crescere di n, e quindi

$$P\left(X \geq \frac{n}{2}\right) \geq 1 - 2^{-n/8}.$$

□

9 Ulteriori teoremi sulle componenti

- Se $p(n-1) < 1$, allora quasi sicuramente tutte le componenti connesse hanno dimensione $O(\log n)$.
- Se $p(n-1) = 1$, allora quasi sicuramente esiste una componente connessa di dimensione approssimativamente $n^{2/3}$.
- Se $p(n-1) > 1$, allora quasi sicuramente esiste una componente connessa gigante $\Omega(n)$, mentre le altre hanno dimensione $O(\log n)$.

Qui "quasi sicuramente" significa che la probabilità dell'evento tende a 1 quando $n \to \infty$.

10 Grado medio dei nodi in $G_{n,p}$

Sia δ_i la variabile aleatoria che rappresenta il grado del nodo i.

Sia a_{ij} la variabile aleatoria che vale 1 se il nodo i è connesso con il nodo j, 0 altrimenti.

Si ha:

$$\delta_i = \sum_{j \in [n]-i} a_{ij}$$

Il valore atteso del grado è:

$$E[\delta_i] = \sum_{j \in [n]-i} E[a_{ij}] = \sum_{j \in [n]-i} p = (n-1)p$$

Se p è costante, il grado cresce linearmente con n, il che non è realistico per reti sociali.

Per modelli più significativi si prende $p = \frac{\lambda}{n}$ con $\lambda > 0$ costante.

11 Distribuzione del grado

La probabilità che un nodo i abbia grado k è:

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

Sostituendo $p = \frac{\lambda}{n}$ e per n grande, si ottiene una distribuzione di Poisson approssimata:

$$P(\delta_i = k) \approx e^{-\lambda} \frac{\lambda^k}{k!}$$

12 Fractions dei nodi di grado k

Indichiamo con F_k la frazione di nodi con grado k:

$$F_k = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\delta_i = k}$$

Il valore atteso è:

$$E[F_k] = P(\delta_i = k) \approx e^{-\lambda} \frac{\lambda^k}{k!}$$

La probabilità decresce esponenzialmente in k, cioè la distribuzione è ben concentrata.

13 Risultato tecnico utile

$$1 - x \leq e^{-x} \quad \text{per ogni } x \in \mathbb{R}$$

Questa disuguaglianza è usata ripetutamente nelle stime probabilistiche del modello.