Introduzione al corso di Analisi di Reti

Con qualche cenno al modello di Erdös-Renyi

Premessa

- In genere, quando vengono resi disponibili agli studenti i lucidi delle lezioni, gli studenti si limitano a studiare sui lucidi.
- Questo comportamento risulta sempre (e necessariamente) in una preparazione superficiale e insufficiente
- Perciò, il libro di testo (e, in alcuni casi, qualche dispensa che renderò disponibile on-line) è un necessario complemento al materiale contenuto in questi lucidi al fine di conseguire una preparazione adeguata
- Nei lucidi indicherò sempre la porzione di testo (o la dispensa) cui far riferimento per l'argomento trattato
  - con qualche eccezione: ad esempio, il modello di Erdös-Renyi, di cui ci occupiamo in questa lezione, viene trattato solo in questi lucidi

Analisi di Reti

- Obiettivo del corso: partiamo dal nome
- Analisi di **Reti**
  - ci occuperemo di reti – nell'accezione del termine più ampia possibile
- **Analisi** di Reti
  - e le reti le analizzeremo
  - da molti punti vista: **prestazioni**, **struttura**, **utilizzo**…
  - e utilizzando **tecniche** prese in prestito da numerose discipline
  - e riservando attenzione a queste tecniche: ossia, *gli argomenti che tratteremo saranno anche uno spunto per studiare le tecniche utilizzate per analizzarli*
- I contenuti del corso sono una sintesi fra
  - Matematica/informatica (modelli, analisi, algoritmi, complessità)
  - Economia (relazioni come incentivo/disincentivo a comportamenti)
  - Scienze sociali (studio di strutture e interazioni caratteristiche di gruppi e popolazioni)

Analisi di Reti

- Obiettivo del corso: partiamo dal nome
- Dunque, Analisi di Reti (mod. 1) non è un corso a carattere puramente informatico
  - ad esempio, non è un corso di Algoritmi
- Analisi di Reti (mod. 1) è un corso che si occupa di *modelli* e della loro analisi
  - cercheremo di generare reti capaci di modellare talune caratteristiche esibite dalle reti reali
    - e dimostreremo che le reti generate nel modo che abbiamo pensato esibiscono effettivamente quelle caratteristiche
  - modelleremo alcune caratteristiche delle reti mediante definizioni formali
    - e studieremo fino a che punto i modelli sono rappresentativi e/o utilizzabili
  - modelleremo formalmente alcuni dei problemi che si presentano nella gestione delle reti
    - e, neanche a dirlo, ne cercheremo le soluzioni

Reti

- Genericamente parlando, una rete è uno schema di interconnessione fra un insieme di entità
- Dipendentemente dal tipo di entità parliamo di reti fisiche, reti sociali, reti di informazioni…
- L'idea di rete alla base di questo corso: *ampia popolazione che reagisce alle azioni dei singoli*
- Ciò che ci interessa è studiare il comportamento *aggregato* di gruppi di individui:
  - Come la presenza di legami influisce sul comportamento dei singoli individui (effetti informativi, fenomeni di diffusione, ricerca decentralizzata di percorsi brevi)
  - Come la presenza di legami modifica la struttura stessa della rete (stabilità, fenomeno rich get richer)
  - come la struttura dell'insieme dei legami permette di desumere informazioni (web-search, sistemi di voto)
  - Una descrizione di tutto ciò lo trovate nel Cap. 1, pag. 1-17

Struttura di una rete

- È difficile rappresentare e studiare puntualmente reti di milioni di individui
- si possono però analizzare proprietà "globali" di una rete di grandi dimensioni
  - se la rete contiene Componenti Giganti:
    - esperimento di Leskovec e Horvitz su Instant Messenger mostra che il grafo delle conversazioni bidirezionali in un mese ha una componente di 200 milioni di utenti su 240 milioni totali
  - all'interno di una componente connessa può avere interesse ricercare porzioni *densamente* connesse
  - si può studiare se la rete presenta una struttura centro / periferia
  - si può studiare il ruolo dei nodi che costituiscono una porzione densamente connessa, suddividendoli in entità centrali / periferiche
  - …

Struttura di una rete

- Talvolta, lo studio dei fenomeni lo porteremo avanti *a livello di popolazione* - senza considerare i singoli individui
  - quando consideriamo il fenomeno rich get richer, ad esempio, studiamo quale è, *mediamente*, la frazione di individui che ha un elevato grado di popolarità
  - non studiamo, individuo per individuo, qual è la sua popolarità!
- Talvolta per comprendere altri fenomeni, occorrerà considerare la struttura fisica della rete
  - ad esempio, per comprendere il ruolo di una certa relazione all'interno di una data rete, dobbiamo studiare precisamente la topologia di quella rete
    - come accade nello studio dei fenomeni di diffusione
  - relazione per relazione!

Struttura di una rete

- In ogni caso, poiché una rete è un insieme di individui e delle relazioni che intercorrono fra essi
- neanche a dirlo, una rete viene modellata attraverso un grafo!
- E per studiare i fenomeni che avvengono all'interno di una rete utilizzeremo tutte le nozioni che già conosciamo di teoria dei grafi
  - grafi non orientati / orientati
  - percorsi
  - spanning trees
  - componenti connesse
  - diametro
  - BFS …
- Tutto ciò è descritto nel capitolo 2 del testo (argomenti che assumo noti)

C’è rete e rete

- Sì, ma per studiare i fenomeni che avvengono all’interno di *una* rete, *quella* rete bisogna averla sotto gli occhi!
- Così, per esempio, qualcuno ci mostra una certa rete, noi la modelliamo tramite un grafo, e poi studiamo le sue proprietà
  - se la rete contiene Componenti Giganti
  - possiamo studiare la funzione che, in quel grafo, esprime il numero di nodi che hanno un certo grado
  - possiamo studiare il diametro di quel grafo
  - …
- Però, potremmo essere interessati a studiare le stesse questioni da un altro punto di vista…

C'è rete e rete

- Però, potremmo essere interessati a studiare le stesse questioni da un altro punto di vista:
  - qual è, *mediamente*, il diametro di una rete sociale, espresso in funzione del numero dei nodi?
  - in funzione del grado, qual è, *mediamente*, la frazione del numero di nodi che ha un certo grado in una rete sociale?
- Per occuparci di queste questioni occorre considerare tante (ma tante) reti
- E queste reti, dove le prendiamo????
- Sì perché ottenere i dati di una rete vera non è una cosa facile facile
  - i dati, chi ce li ha se li tiene!
- E poi, comunque, di reti vere, in circolazione, ce ne sono un numero limitato
  - comunque non abbastanza per un'indagine statistica
- E allora?! Vorrà dire che le reti ce le dovremo inventare…

Modelli generativi di grafi casuali

- Ma cosa vuol dire "inventare" una rete?
- Beh, semplice: generare un grafo in modo casuale
  - ossia, un grafo in cui gli archi fra i nodi sono scelti sulla base di un evento aleatorio
  - per esempio. sulla base del lancio di una moneta
- Sono stati proposti molti modelli per generare grafi casuali
  - ossia, regole probabilistiche che permettono di connettere nodi
  - e noi ne studieremo quattro
- Naturalmente, **alcuni di questi modelli riprodurranno taluni fenomeni che sono stati osservati nelle reti reali, ma non tutti!**
- E noi cercheremo di capire *quali modelli* utilizzare per riprodurre *quale fenomeno*…

Il modello di Erdös-Renyi

- Fissiamo $n \in \mathbb{N}$ e $p \in [0,1]$
- A partire da n e p costruiamo un grafo nel modo seguente
  - l'insieme dei nodi del grafo è $[n] = \{1,2, \dots , n\}$
  - per ogni coppia di elementi distinti i e j in [n]: con probabilità p viene inserito nel grafo l'arco (i , j)
- Il grafo così costruito è un evento aleatorio che indichiamo con $\mathbf{G_{n,p}}$
- **$G_{n,p}$ è un grafo aleatorio**
- del quale ora studieremo alcune caratteristiche
  - esistenza di una componente gigante
  - grado dei nodi
- neanche a dirlo, utilizzando il calcolo delle probabilità!

Il modello di Erdös-Renyi

- Osserviamo, innanzi tutto, che, fissato il numero n dei nodi, *al variare di p* si otterranno grafi con caratteristiche molto differenti
  - quando p = 0 sarà possibile ottenere un unico grafo $G_{n,0}$: quello che non contiene alcun arco!
  - Analogamente, quando p = 1 sarà possibile ottenere un unico grafo $G_{n,1}$: il grafo completo su n nodi!
  - In generale, $G_{n,p}$ conterrà *mediamente* tanti più archi quanto più p si avvicina a 1
- In particolare, *mediamente*, un nodo avrà tanti più vicini quanto più p si avvicina a 1
- E, *mediamente*, le componenti connesse di $G_{n,p}$ saranno tanto più grandi quanto più p si avvicina a 1
- Non resta che quantificare!

Il modello di Erdös-Renyi – componenti giganti

- Una ***componente gigante*** in un grafo è una componente connessa che contiene <u>una frazione del numero dei nodi</u>
- Molte reti reali contengono una componente gigante
  - questo è vero nelle reti sociali - come ha mostrato l'esperimento di Leskovec e Horvitz nella di conversazioni su Instant Messenger
  - e, come vedremo più avanti, è vero anche nella rete del Web – i cui nodi sono le pagine presenti nel web e gli archi gli hyperlink fra esse
- Cominciamo col domandarci: il modello di Erdos-Renyi riesce a rappresentare questa caratteristica di molte reti reali?
- Più precisamente: esistono valori del parametro p per i quali $G_{n,p}$ contiene componenti giganti?
- Come primo risultato dimostreremo che
- ***se p >*** $\frac{\ln 64}{n}$ ***allora con alta probabilità*** $G_{n,p}$ ***contiene una componente connessa costituita da almeno metà dei suoi nodi***

*

Il modello di Erdös-Renyi – componenti giganti

- se p > $\frac{\ln 64}{n}$ allora con alta probabilità $G_{n,p}$ contiene una componente connessa costituita da almeno metà dei suoi nodi
- Chiariamo:
  - intanto, ***con alta probabilità*** significa **con probabilità proporzionale *almeno* a** $\mathbf{(1 - \frac{b}{n^c})}$ per qualche coppia di costanti *positive* b e c
    - cioè, al crescere di n la probabilità tende *velocemente* a 1
  - poi, definiamo **X** come la variabile aleatoria corrispondente al numero di nodi nella più grande componente connessa di $G_{n,p}$
- allora, il teorema è così enunciato formalmente:

<u>Teorema</u>: se p > $\frac{\ln 64}{n}$ allora P(X ≥ $\frac{n}{2}$)≥ 1- $2^{-\frac{n}{8}}$

- Per dimostrare il teorema abbiamo bisogno di un lemma
  - e della notazione [n] = {1, 2, … , n}

Il modello di Erdös-Renyi – componenti giganti

- Per dimostrare il teorema abbiamo bisogno di un lemma:

  Lemma: se $X < n/2$ allora esiste un insieme $A \subset [n]$ tale che $n/4 \leq |A| < 3n/4$ e non esistono archi fra i nodi in A e i nodi in [n]-A

- **Dimostrazione**: siano $C_1, C_2, \ldots, C_k$ tutte le componenti connesse di $G_{n,p}$ e
  - assumiamo che siano ordinate per cardinalità non decrescente, ossia
    $$|C_1| \leq |C_2| \leq \ldots \leq |C_k|$$
  - poiché $X < n/2$, allora $|C_i| < n/2$ per ogni $i=1, \ldots, k$
  - scegliamo un indice h tale che:
    $$|C_1| + |C_2| + \ldots + |C_{h-1}| < n/4 \quad \text{e}$$
    $$|C_1| + |C_2| + \ldots + |C_{h-1}| + |C_h| \geq n/4$$
    - **$h < k$**: infatti, poiché $|C_k| < n/2$, se fosse $h=k$ risulterebbe
      $$|C_1| + |C_2| + \ldots + |C_{k-1}| + |C_k| < n/4 + n/2 < n$$
  - **Scegliamo $A = C_1 \cup C_2 \cup \ldots \cup C_h$**
    - Allora, $A \neq \emptyset$ e $[n] - A = C_{h+1} \cup \ldots \cup C_k \neq \emptyset$

Il modello di Erdös-Renyi – componenti giganti

- Per dimostrare il teorema abbiamo bisogno di un lemma:

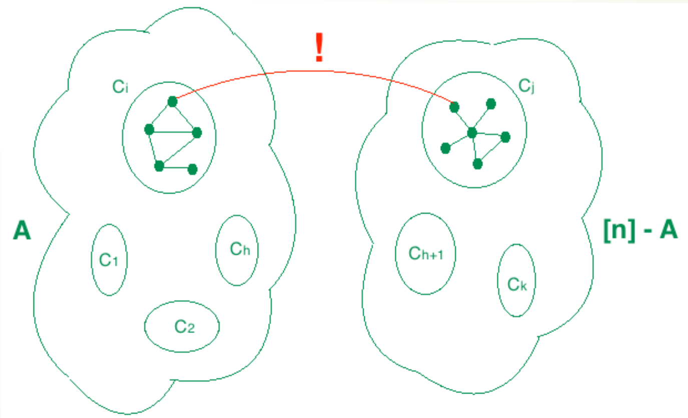
Lemma: se $X < n/2$ allora esiste un insieme $A \subset [n]$ tale che $n/4 \leq |A| < 3n/4$ e non esistono archi fra i nodi in A e i nodi in $[n]-A$

- **Dimostrazione**: siano $C_1, C_2, \ldots, C_k$ tutte le componenti connesse di $G_{n,p}$ con
$$|C_1| \leq |C_2| \leq \ldots \leq |C_k|$$
  - poiché $X < n/2$, allora $|C_i| < n/2$ per ogni $i=0,1, \ldots, k$
  - scegliamo un indice h (< k) tale che:
$$|C_1| + |C_2| + \ldots + |C_{h-1}| < n/4 \quad \text{e} \quad |C_1| + |C_2| + \ldots + |C_{h-1}| + |\mathbf{C_h}| \geq \mathbf{n/4}$$
  - Scegliamo $\mathbf{A = C_1 \cup C_2 \cup \ldots \cup C_h}$
  - Allora,
    - $|A| \geq n/4$ per costruzione,
    - e, come abbiamo visto, $|A| = (|C_1| + \ldots + |C_{h-1}|) + |C_h| < n/4 + n/2 = 3n/4$

Il modello di Erdös-Renyi – componenti giganti

- **Lemma: se $X < n/2$ allora esiste un insieme $A \subset [n]$ tale che $n/4 \leq |A| < 3n/4$ e non esistono archi fra i nodi in A e i nodi in [n]-A**
- **Dimostrazione:** $C_1, C_2, \dots, C_k$ sono le componenti connesse di $G_{n,p}$ con $|C_1| \leq |C_2| \leq \dots \leq |C_k|$
  - Abbiamo scelto $h < k$ tale che $|C_1| + |C_2| + \dots + |C_{h-1}| < n/4$ e $|C_1| + |C_2| + \dots + |C_{h-1}| + |\mathbf{C_h}| \geq \mathbf{n/4}$
  - E abbiamo posto $\mathbf{A = C_1 \cup C_2 \cup \dots \cup C_h}$ e, dunque, $\mathbf{[n]\text{-}A = C_{h+1} \cup \dots \cup C_k}$
  - per costruzione, poichè $C_1, C_2, \dots, C_k$ sono tutte le componenti connesse di $G_{n,p}$, non ci sono archi fra A e [n]-A

    altrimenti, se ci fosse un arco fra C_i e C_j, con $i \leq h$ e $j > h$, allora $C_i \cup C_j$ sarebbe un'unica componente connessa!



Il modello di Erdös-Renyi – componenti giganti

- **Lemma: se $X < n/2$ allora esiste un insieme $A \subset [n]$ tale che $n/4 \leq |A| < 3n/4$ e non esistono archi fra i nodi in A e i nodi in [n]-A**
- **Dimostrazione:** $C_1, C_2, \dots, C_k$ sono le componenti connesse di $G_{n,p}$ con $|C_1| \leq |C_2| \leq \dots \leq |C_k|$
  - Abbiamo scelto $h < k$ tale che $|C_1| + |C_2| + \dots + |C_{h-1}| < n/4$ e $|C_1| + |C_2| + \dots + |C_{h-1}| + |C_h| \geq n/4$
  - E abbiamo posto $A = C_1 \cup C_2 \cup \dots \cup C_h$ e, dunque, $[n]\text{-}A = C_{h+1} \cup \dots \cup C_k$
  - Riassumendo:
    - 1) $|A| \geq n/4$ per costruzione, e
    - 2) $|A| < 3n/4$
    - 3) **non ci sono archi fra A e [n]-A**
- Chiamiamo ***buono*** l'insieme A individuato dal lemma
- **OSSERVAZIONE**: il lemma ci assicura che ***la probabilità che la più grande componente connessa di $G_{n,p}$ contenga meno di $n/2$ nodi è minore o uguale alla probabilità che $G_{n,p}$ contenga un insieme di nodi buono***
  - perché, dati i due eventi e_1 ed e_2, se **e_1 implica e_2** allora $P(e_1) \leq P(e_2)$

Il modello di Erdős-Renyi – componenti giganti

- **<u>Teorema</u>: se p > $\frac{\ln 64}{n}$ allora P(X ≥ $\frac{n}{2}$) ≥ 1- 2 $^{-n/8}$**
- **Dimostrazione**: calcoliamo **P(X <** $\frac{n}{2}$ **)** , ossia la probabilità che la massima componente connessa in $G_{n,p}$ contenga meno di n/2 nodi
  - che è l'evento complementare dell'evento "**X ≥** $\frac{n}{2}$" che ci interessa
- in virtù dell'OSSERVAZIONE (a pag. precedente):

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

- Allora:

$P(X < \frac{n}{2}) \leq P(\exists\, A \subset [n]$: A è buono) *da cui, ricordando la definizione di insieme buono*

$$= P\left(\bigcup_{A \subset [n]\ e\ \frac{n}{4} \leq |A| < \frac{3n}{4}} [\text{ non ci sono archi fra A e [n]-A) }] \right)$$

applichiamo ora lo Union Bound: la probabilità dell'unione di eventi è ≤ della somma delle probabilità degli eventi

$$\leq \sum_{A \subset [n]\ e\ \frac{n}{4} \leq |A| < \frac{3n}{4}} P(\text{non ci sono archi fra A e [n]-A})$$

Il modello di Erdös-Renyi – componenti giganti

- **Teorema: se p > $\frac{ln\,64}{n}$ allora P(X ≥ $\frac{n}{2}$) ≥ 1- 2 $^{-n/8}$**
- **Dimostrazione**: calcoliamo **P(X < $\frac{n}{2}$) :**

$$P(X < \tfrac{n}{2}) \leq \sum_{A \subset [n]\ e\ \frac{n}{4} \leq |A| < \frac{3n}{4}} P(\text{non ci sono archi fra A e [n]-A})$$

$$= \sum_{A \subset [n]\ e\ \frac{n}{4} \leq |A| \leq \frac{3n}{4}} (1-p)^{|A|(n-|A|)}$$

perché il numero di archi possibili fra A e [n]-A è |A|·|[n]-A| e la probabilità che un arco possibile non sia un arco è (1-p)

$$\leq \sum_{A \subset [n]\ e\ \frac{n}{4} \leq |A| \leq \frac{3n}{4}} (1-p)^{3n^2/16}$$

perchè (1-p) < 1 (ossia, $(1-p)^z$ è massimo quando z è minimo) e |A|(n-|A|) è minimo per |A| = n/4 (o, equivalentemente, |A|= 3n/4)

Il modello di Erdös-Renyi – componenti giganti

- **Teorema**: **se p > $\frac{\ln 64}{n}$ allora P(X ≥ $\frac{n}{2}$) ≥ 1- 2 $^{-n/8}$**
- **Dimostrazione**: calcoliamo P(X < $\frac{n}{2}$):

$$P(X < \tfrac{n}{2}) \leq \sum_{A \subset [n] \text{ e } \frac{n}{4} \leq |A| \leq \frac{3n}{4}} (1-p)^{3n^2/16}$$

$$\leq 2^n (1-p)^{3n^2/16}$$

perché [n] contiene 2^n sottoinsiemi

$$< 2^n\left(1 - \frac{\ln 64}{n}\right)^{\frac{3n^2}{16}}$$

(e qui si usa l'ipotesi e c'è un minore stretto!)

$$= 2^n \left[\left(\mathbf{1} - \frac{\ln 64}{n}\right)^{n}\right]^{\frac{3n}{16}}$$

ora, moltiplichiamo e dividiamo l'esponente n rosso per – ln 64

$$= 2^n \left[\left(\mathbf{1} - \frac{\ln 64}{n}\right)^{\frac{-n}{\ln 64}(-\ln 64)}\right]^{\frac{3n}{16}}$$

Il modello di Erdös-Renyi – componenti giganti

- **Teorema: se p > $\frac{\ln 64}{n}$ allora P(X ≥ $\frac{n}{2}$)≥ 1- 2 $^{-n/8}$**
- **Dimostrazione**: $P(X < \frac{n}{2}) < 2^n \left[\left(1 - \frac{\ln 64}{n} \right)^{\frac{-n}{\ln 64}(-\ln 64)} \right]^{\frac{3n}{16}}$
  - Poiché $\lim_{n \to \infty} \left(1 - \frac{\ln 64}{n} \right)^{\frac{-n}{\ln 64}} = e$, allora per n sufficientemente grande

    $$\left(1 - \frac{\ln 64}{n} \right)^{\frac{-n}{\ln 64}(-\ln 64)} \approx e^{-\ln 64} = 64^{-1}$$

  - e dunque

    $$P(X < \frac{n}{2}) < 2^n [64^{-1}]^{3n/16} = 2^n \left[2^{-6} \right]^{3n/16} = 2^n 2^{-18n/16} = 2^{-n/8}$$

  - da cui segue il teorema

Il modello di Erdös-Renyi – componenti giganti

- **Teorema**: se p > $\frac{\ln 64}{n}$ allora P(X ≥ $\frac{n}{2}$) ≥ 1- 2 $^{-n/8}$
- Accanto al precedente teorema si può dimostrrare anche che:

  Teorema:

  1) se p(n-1) < 1 allora quasi sicuramente tutte le componenti connesse di $G_{n,p}$ hanno O(log n) nodi

  2) se p(n-1) = 1 allora quasi sicuramente $G_{n,p}$ ha una componente connessa di ≈ n $^{2/3}$ nodi

  3) se p(n-1) > 1 allora quasi sicuramente $G_{n,p}$ ha una componente connessa di Ω(n) nodi e tutte le altre componenti connesse hanno O(log n) nodi

- ***"quasi sicuramente"*** significa che, al tendere di n all'infinito la probabilità dell'evento tende a 1
  - un po' meno che con alta probabilità
- In conclusione, la presenza di componenti giganti dipende dal prodotto p(n-1)
  - ma cosa rappresenta p(n-1)?

Il modello di Erdös-Renyi – grado dei nodi

- In conclusione, la presenza di componenti giganti dipende dal prodotto p(n-1)
  - ma cosa rappresenta p(n-1)?
- Per i ∈ [n], indichiamo con δ_i **la variabile aleatoria che esprime il grado del nodo i,**
- Per i ∈ [n] e j ∈ [n]-{i}, indichiamo con a_{ij} la variabile aleatoria che vale 1 se il nodo i è adiacente al nodo j vale 0 altrimenti
- Allora, poiché $\delta_i = \sum_{j \in [n] - \{i\}} a_{ij}$, il valore atteso del grado di un nodo è

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

$$\sum_{j \in [n] - \{i\}} [1 \cdot p + 0 \cdot (1-p)] = \sum_{j \in [n] - \{i\}} p \mathbf{= (n-1)p}$$

- Questo significa che, se p è una costante, il grado dei nodi, in media, cresce linearmente con il numero dei nodi

Il modello di Erdös-Renyi – grado dei nodi

- Per i ∈ [n], se indichiamo con **δ_i la variabile aleatoria che esprime il grado del nodo i,** abbiamo che il valore atteso del grado di un nodo è

  $E[\delta_i] = (n-1)p$

- Questo significa che, se p è una costante, il grado dei nodi, in media, cresce linearmente con il numero dei nodi
- Cerchiamo ora di capire se $G_{n,p}$ ben si presta a descrivere una rete sociale
  - costituita da tanti, ma proprio tanti tanti, individui
- Se il grado medio di un nodo è (n-1)p , e p è un valore costante, allora, mediamente, il numero di contatti di un individuo in una rete sociale è proporzionale agli individui che costituiscono la rete sociale!
- Che non è propriamente ragionevole!
- Per questa ragione, al fine di modellare significativamente reti reali di grandi dimensioni è opportuno che p sia una funzione decrescente di n
  - del tipo $p = p(n) = \frac{\lambda}{n}$

Il modello di Erdös-Renyi - grado dei nodi

- Scegliamo, dunque, $p = p(n) = \frac{\lambda}{n}$ per qualche costante positiva λ

- Fissato un intero $k < n$, vogliamo ora calcolare con quale probabilità un nodo in $G_{n,p}$ ha grado k

- Sia $i \in [n]$: la probabilità che il nodo i abbia grado k è la probabilità che *esattamente* k altri nodi siano adiacenti a i
  - il numero di possibili k-uple di nodi scelti nell'insieme $[n] - \{ i \}$ è $\binom{n-1}{k}$
  - la probabilità che vi sia un arco fra i e *ciascuno* dei nodi della k-upla è p^k
  - la probabilità che **non** vi sia un arco fra i e *ciascuno* dei nodi non contenuto nellla k-upla è $(1-p)^{n-1-k}$

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

Il modello di Erdös-Renyi - grado dei nodi

- Scegliamo, dunque, p = p(n) = $\frac{\lambda}{n}$ per qualche costante positiva λ
- Fissato un intero $k \leq n$, vogliamo ora calcolare con quale probabilità un nodo in $G_{n,p}$ ha grado k: sia $i \in [n]$
- $P(\delta_i = k) = \binom{n-1}{k} p^k (1-p)^{n-k-1} =$ *da cui, sostituendo* $p = \frac{\lambda}{n}$

$$= \binom{n-1}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k-1} =$$

esplicitando il coefficiente binomiale

$$= \frac{(n-1)\cdot(n.2)\cdot \ldots \cdot(n-k)}{k!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k-1}$$

ora: poiché $(1-x) < e^{-x}$ *per ogni* $0 < x < 1$ *(vedi lucido ''Risultato tecnico" al termine della prova), allora* $\left(1 - \frac{\lambda}{n}\right)^{n-k-1} < e^{\frac{-\lambda\,(n-k-1)}{n}}$

$$< \frac{(n-1)\cdot(n-2)\cdot \ldots \cdot(n-k)}{k!} \frac{\lambda^k}{n^k} e^{\frac{-\lambda\,(n-k-1)}{n}}$$

Il modello di Erdös-Renyi - grado dei nodi

- Scegliamo, dunque, $p = p(n) = \frac{\lambda}{n}$ per qualche costante positiva λ
- Fissato un intero $k \leq n$, vogliamo ora calcolare con quale probabilità un nodo in $G_{n,p}$ ha grado k: sia $i \in [n]$
- $P(\delta_i = k) = \binom{n-1}{k} p^k (1-p)^{n-k-1} < \frac{(n-1)\cdot(n-2)\cdot \dots \cdot(n-k)}{k!} \frac{\lambda^k}{n^k} e^{\frac{-\lambda\,(n-k-1)}{n}}$

$$< \frac{n^k}{k!} \frac{\lambda^k}{n^k} e^{\frac{-\lambda\,(n-k-1)}{n}}$$

$$\approx \frac{n^k}{k!} \frac{\lambda^k}{n^k} e^{-\lambda}$$ *per n sufficientemente grande*

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

Il modello di Erdös-Renyi - grado dei nodi

- Scegliamo, dunque, $p = p(n) = \frac{\lambda}{n}$ per qualche costante positiva λ
- $P(\delta_i = k) < \frac{\lambda^k}{k!} e^{-\lambda}$
- e poiché $k! \approx \sqrt{2 \pi k} \left(\frac{k}{e}\right)^k$ *(approssimazione di Stirling)*
- $P(\delta_i = k) < \frac{(\lambda e)^k}{\sqrt{2 \pi k}\, k^k} e^{-\lambda} = \frac{e^{-\lambda}}{\sqrt{2 \pi k}} \left(\frac{\boldsymbol{\lambda} \ \mathbf{e}}{\mathbf{k}}\right)^{\mathbf{k}}$

- Dunque, possiamo concludere che **la probabilità che un generico nodo abbia grado k** decresce *molto velocemente* al crescere di k
- più precisamente, decresce come k^{-k} – ossia, **decresce esponenzialmente in k**

Risultato tecnico

- A questo punto, dimostriamo il risultato tecnico che abbiamo utilizzato
- **<u>Lemma</u>: Per ogni $x \in \mathbb{R}$: $1-x \leq e^{-x}$. Inoltre, se $x \neq 0$ allora $1-x < e^{-x}$.**
- Definiamo la funzione $G(x) = 1-x-e^{-x}$
- Calcoliamo la derivata prima di $G(x)$: $G'(x) = e^{-x} - 1$
- Studiamo il segno di $G'(x)$: $e^{-x} - 1 \geq 0 \ \rightarrow e^{-x} \geq 1 \ \rightarrow e^{-x} \geq e^{0} \ \rightarrow x \leq 0$
- $G'(x) \geq 0$ per $x \leq 0$: allora, $G(x)$ ha un punto di massimo relativo in $x = 0$
- inoltre, essendo l'unico punto in cui la derivata si annulla, $x = 0$ è anche un punto di massimo assoluto
- Poiché $G(0) = 1-0-e^{-0} = 0$, questo implica che
  - $G(x) \leq G(0) = 0$ per ogni $x \in \mathbb{R}$, ossia $1-x \leq e^{-x}$ per ogni $x \in \mathbb{R}$
  - $G(x) < G(0) = 0$ per ogni $x \neq 0$

Il modello di Erdös-Renyi - grado dei nodi

- Scegliamo, dunque, $p = p(n) = \frac{\lambda}{n}$ per qualche costante positiva λ
- allora, $P(\delta_i = k) < \frac{e^{-\lambda}}{\sqrt{2\pi k}} \left(\frac{\lambda e}{k}\right)^k$
- Ci chiediamo, ora, quale sia la frazione del numero di nodi che hanno grado k
- Per rispondere a questa nuova domanda, indichiamo con $\mathbf{F_k}$ la variabile aleatoria che esprime tale frazione
- e indichiamo con $\boldsymbol{\delta_{i\,k}}$ la variabile aleatoria che vale: 1 se $\delta_i = k$, 0 altrimenti

$$\boldsymbol{\delta_{i\,k}} = \begin{cases} 1 & se\ \delta_i = k \\ 0 & altrimenti \end{cases}$$

- allora, $\mathbf{F_k} = \frac{1}{n} \sum_{i \in [n]} \boldsymbol{\delta_{i\,k}}$

Il modello di Erdős-Renyi - grado dei nodi

- Scegliamo, dunque, $p = p(n) = \frac{\lambda}{n}$ per qualche costante positiva λ
- Ci chiediamo, ora, quale sia la frazione del numero di nodi che hanno grado k:
  - indicando con F_k la variabile aleatoria che esprime tale frazione e con δ_{ik} la variabile aleatoria che vale: 1 se $\delta_i = k$, 0 altrimenti, allora, $F_k = \frac{1}{n} \sum_{i \in [n]} \delta_{ik}$
  - e ricordando che $P(\delta_i = k) < \frac{e^{-\lambda}}{\sqrt{2 \pi k}} \left(\frac{\lambda\ e}{k}\right)^k$
- Otteniamo, $\mathbf{E[\,F_k\,]} = E\left[\frac{1}{n} \sum_{i \in [n]} \delta_{ik}\right]$

$$= \frac{1}{n} \sum_{i \in [n]} E[\delta_{ik}]$$

per la linearità del valore atteso

$$= \frac{1}{n} \sum_{i \in [n]} [\, 1 \cdot P(\delta_i = k) + 0 \cdot P(\delta_i \neq k)\,]$$

$$= \frac{1}{n} \sum_{i \in [n]} P(\delta_i = k)$$

$$\boldsymbol{< \frac{e^{-\lambda}}{\sqrt{2 \pi k}} \left(\frac{\lambda\ e}{k}\right)^k}$$

Il modello di Erdös-Renyi - grado dei nodi

- Scegliamo, dunque, $p = p(n) = \frac{\lambda}{n}$ per qualche costante positiva λ
- Ci chiediamo, ora, quale sia la frazione del numero di nodi che hanno grado k:
  - indicando con F_k la variabile aleatoria che esprime tale frazione e con δ_{ik} la variabile aleatoria che vale: 1 se $\delta_i = k$, 0 altrimenti, allora, $F_k = \frac{1}{n}\sum_{i \in [n]} \delta_{ik}$
  - e ricordando che $P(\delta_i = k) < \frac{e^{-\lambda}}{\sqrt{2\pi k}} \left(\frac{\lambda e}{k}\right)^k$
- Otteniamo, $\mathbf{E[F_k] = E[\frac{1}{n}\sum_{i \in [n]} \delta_{ik}] = \frac{1}{n}\sum_{i \in [n]} E[\delta_{ik}] =}$

$$= \frac{1}{n}\sum_{i \in [n]} P(\delta_i = k) < \frac{e^{-\lambda}}{\sqrt{2\pi k}} \left(\frac{\lambda e}{k}\right)^k$$

- Ossia, mediamente, la frazione del numero di nodi che hanno grado k decresce come k^{-k} – ossia, **decresce esponenzialmente in k**
- E vedremo la prossima lezione quanto questo corrisponda a ciò che accade nelle reti reali...