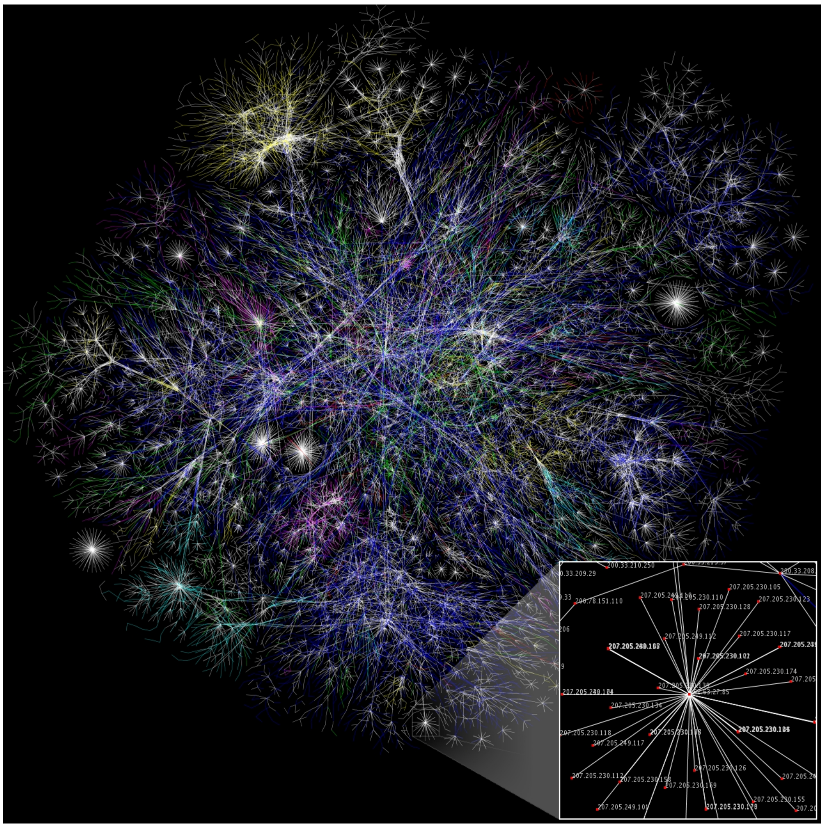
Università di Roma Tre

Geometria per il TFA
grafi e reti

Francesca MEROLA
merola@mat.uniroma3.it
http://ricerca.mat.uniroma3.it/users/merola/

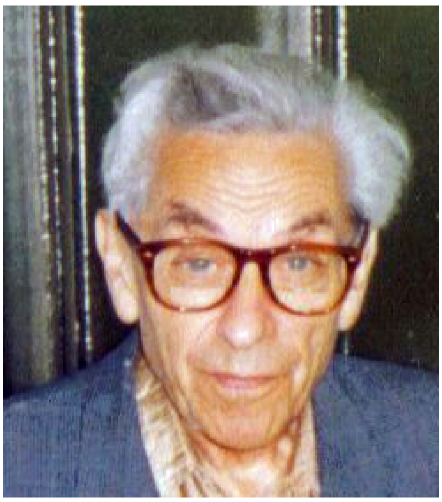


Internet map 1024 by The Opte Project

roadmap

- esempi di grafi e reti sociali e no
- caratteristiche delle reti
- modelli di reti

Paul Erdős (1913–1996)

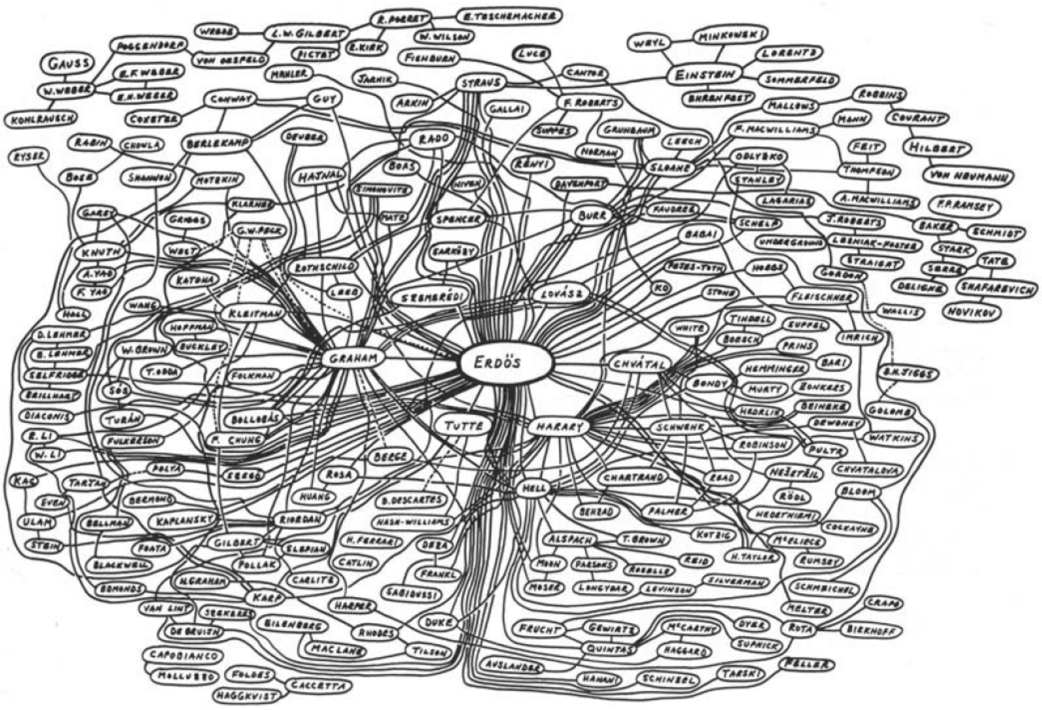


interessi: teoria dei numeri, combinatoria, probablità, analisi. . .
autore (quasi sempre con coautori) di più di 1500 articoli
più di 500 coautori

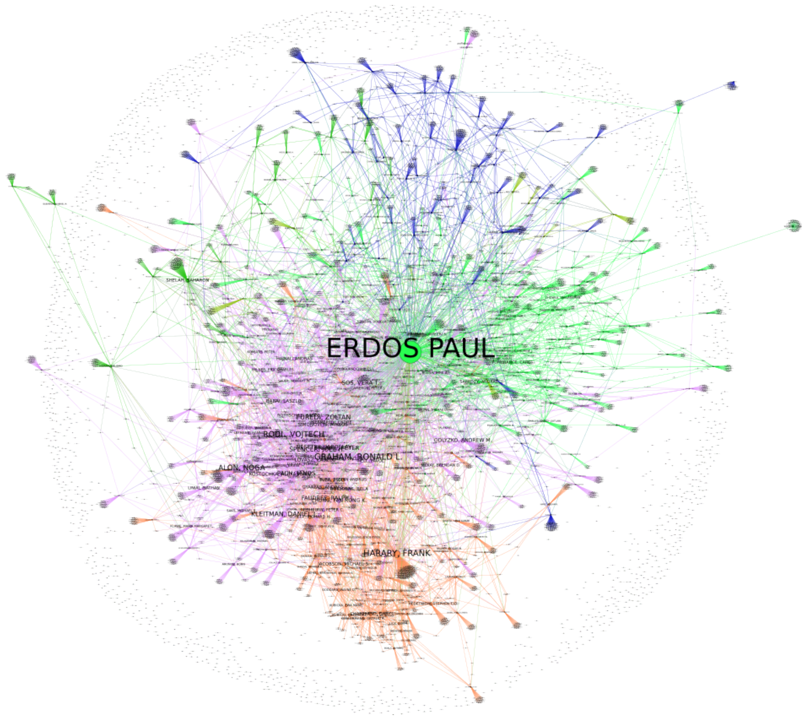
numero di Erdős

- Erdős ha numero di Erdős (NdE) 0
- i coautori di Erdős hanno NdE 1
- i coautori dei coautori di Erdős hanno NdE 2
- e così via
- definiamo un grafo
  - i "matematici" sono i vertici
  - due matematici sono adiacenti se hanno collaborato (scritto un paper di cui sono coautori)
- molti dati su NdE sul sito `http://wwwp.oakland.edu/enp/`

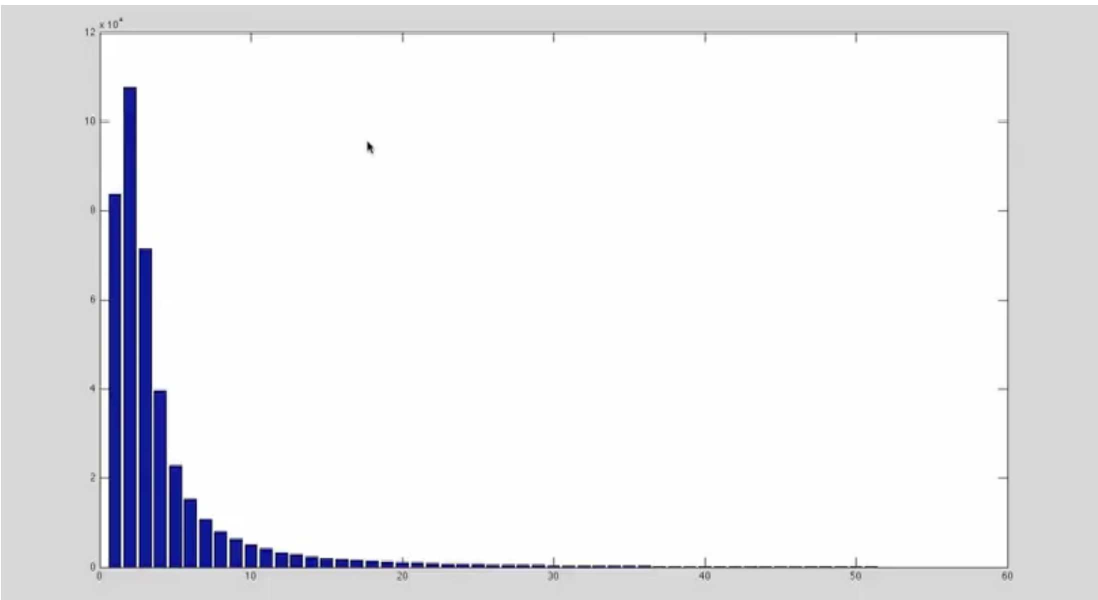
grafo di Erdős - 1



grafo di Erdős - 2



heavy tail



- moltissimi matematici hanno "pochi" coautori
- ma alcuni ne hanno molti - c'è una coda lunga

navigare nelle reti: i sei gradi

- consideriamo un vertice in una rete sociale "grande"
- deve trovare una catena di collegamenti "breve" che lo connetta a un altro vertice
- ha solo una visione locale, non globale della rete
- conosce solo i suoi vicini e qualche informazione su di essi
- come procede?

esperimento di Milgram

- nel 1967 Milgram conduce famosi esperimenti per studiare la rete sociale degli USA
- seleziona un insieme di partenza di circa 260 persone in Nebraska e in Kansas
- e un piccolo insieme di destinatari a Boston
- ogni nodo di partenza riceve informazioni sul suo destinatario - e una lettera da consegnarli
- se conosce direttamente il destinatario gliela invia
- altrimenti invia la lettera a qualcuno che conosce personalmente che possa essere più vicino al destinatario
- in quanti passi si raggiunge l'obiettivo?

esperimento di Milgram - 2

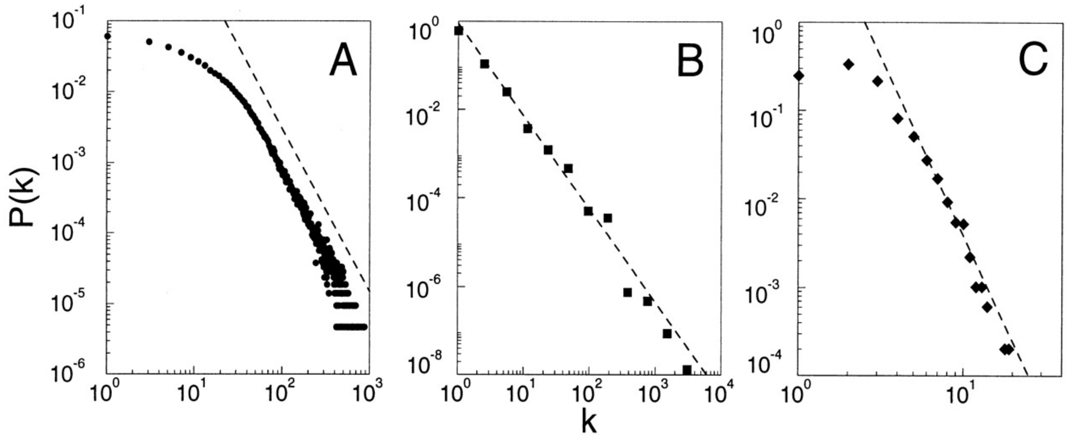
- la maggior parte delle lettere non sono mai arrivate
- quelle che hanno raggiunto il destinatario lo hanno fatto in media in poco meno di sei passi
- inoltre in una versione dell'esperimento, più della metà delle lettere che hanno raggiunto un destinatario
- lo hanno fatto passando per la stessa persona al penultimo passo
- hanno individuato un hub - una persona con molte connessioni
  - presenza di distribuzione con heavy tail
- l'esperimento è stato replicato recentemente su scala più larga via email
- risultati analoghi (Columbia small world project, ca 2003, Watts et al)

caratteristiche delle reti "sociali"

- la distribuzione dei gradi segue una legge di potenza (power law - heavy tail)
- distanza media piccola (i sei gradi di separazione)
- alto coefficiente di clustering (i miei amici tendono a essere amici fra loro)

esempi di legge di potenza (power law)

$Pr(x) = x^{-\beta}$ su scala log log



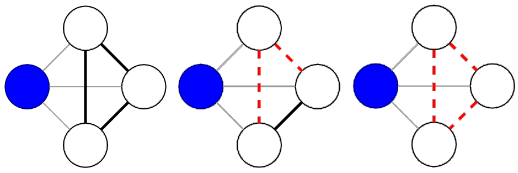
Barabási e Albert, Science, 1999
A rete collaborazione fra attori, ca 200000 vertici, $\beta = 2,3$
B WWW, spigoli = in o out-link, ca 300000 vertici $\beta = 2,1$
C rete elettrica USA, ca 5000 vertici $\beta = 4$

distanza media piccola

- in un grafo connesso con n vertici
- distanza media $\Delta = \frac{2}{n(n-1)} \sum_{u,v \in V} d(u,v)$
- vale 1 per il grafo completo
- è lineare in n per il path (caso peggiore)
- piccola = costante oppure cresce molto lentamente al crescere di n (come $\log\log(n)$)
- esempi dal "mondo reale"
  - i sei gradi dell'esperimento di Milgram
  - Lescovec and Horvitz, 2008: Microsoft Messenger network $\Delta \sim 6,5$, n ca 180M
  - Backstrom et al., 2012: Facebook social graph $\Delta \sim 5$ n ca 721M

coefficiente di clustering

- il coefficiente di clustering locale di un vertice v
- stima quanti vicini di v sono vicini fra loro
- se v ha grado k
- $c(v) = \frac{\#\text{ spigoli che connettono due vicini di } v}{k(k-1)/2}$



c = 1 c = 1/3 c = 0

- il CdC globale $c(G)$ di un grafo G è la media dei CdC dei suoi vertici

coefficiente di clustering "alto"

- cosa vuol dire alto clustering
- voglio confrontarlo con la densità degli spigoli
- $ED = \frac{|E|}{n(n-1)/2} = p$
- alto clustering se $c(G) >> ED$
- mondo reale (Watts 2003)
  - network attori: $c(G) = 0,79$, $ED = 0,00027$
  - rete neuronale del C. elegans $c(G) = 0,28$, $ED = 0,05$
  - rete elettrica USA : $c(G) = 0,080$, $ED = 0,005$

modelli di formazione/crescita reti

- discuteremo tre modelli
- grafi aleatori (Erdős, Rényi)
- il modello del mondo piccolo (Watts e Strogatz)
- preferential attachment (Barabási e Albert)

Erdős e Rényi

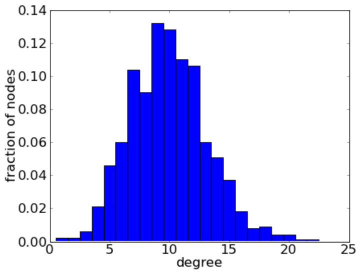
- grafi aleatori (random graphs)
- n vertici inizialmente sconnessi
- si seleziona a caso una coppia di vertici e si aggiunge il relativo spigolo
- si itera questo procedimento
- un modello molto semplice, e poco realistico nel caso delle reti sociali
  - le amicizie non si formano in modo casuale

soglia

- alcune proprietà dei grafi aleatori presentano una soglia
- per esempio la formazione di una componente gigante
- per l'ED al di sotto di un certo valore (qui $1/n$) la probabilità che esista un GC va a zero al crescere di n
- appena l'ED supera il valore soglia si forma una GC (la probabilità che esista tende a uno)
- ogni proprietà monotona delle reti presenta una soglia

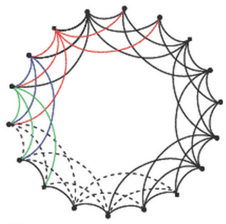
grafi aleatori come modello di reti

- diametro piccolo ok
- clustering alto NO il coefficiente di clustering è pari all'ED
- heavy tail NO la distribuzione dei gradi è una distribuzione di Poisson

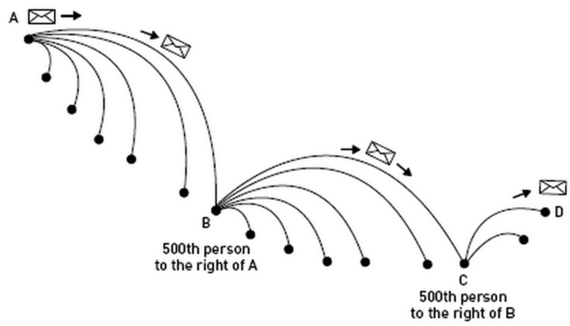


modello di Watts-Strogatz

- questo modello è costruito per avere alto clustering
- si parte da un ciclo con n vertici
- in cui ogni vertice inoltre connesso ai vertici a distanza $\leq k$

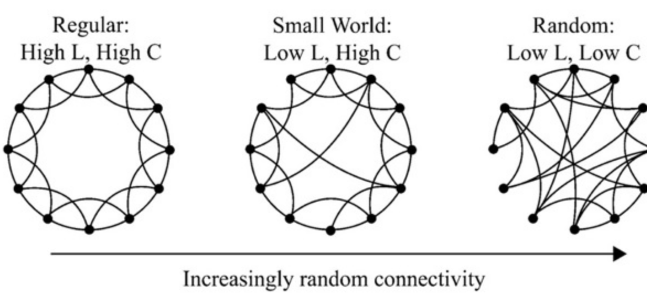


- questo grafo ha alto clustering ma anche diametro alto!



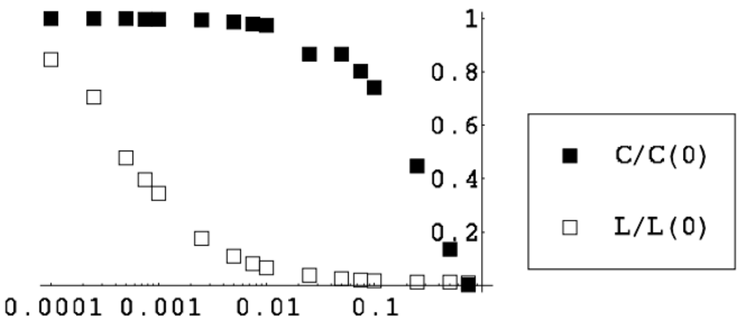
modello di Watts-Strogatz-2

- l'idea è di far diventare alcune "connessioni locali" connessioni casuali
- fissato un vertice, considero i suoi vicini $x_1, x_2, \ldots x_{2k}$
- cambio ognuno di questi con un vertice casuale con probabilità β



modello di Watts-Strogatz-3

- al crescere di β
- la distanza media si abbassa immediatamente
- anche il coefficiente di clustering, ma lentamente
- in un opportuno intervallo ho d.m. bassa e alto clustering



Watts-Strogatz

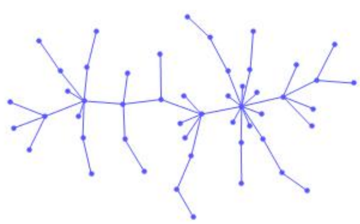
- diametro piccolo ok
- clustering alto ok
- heavy tail NO

the rich get richer

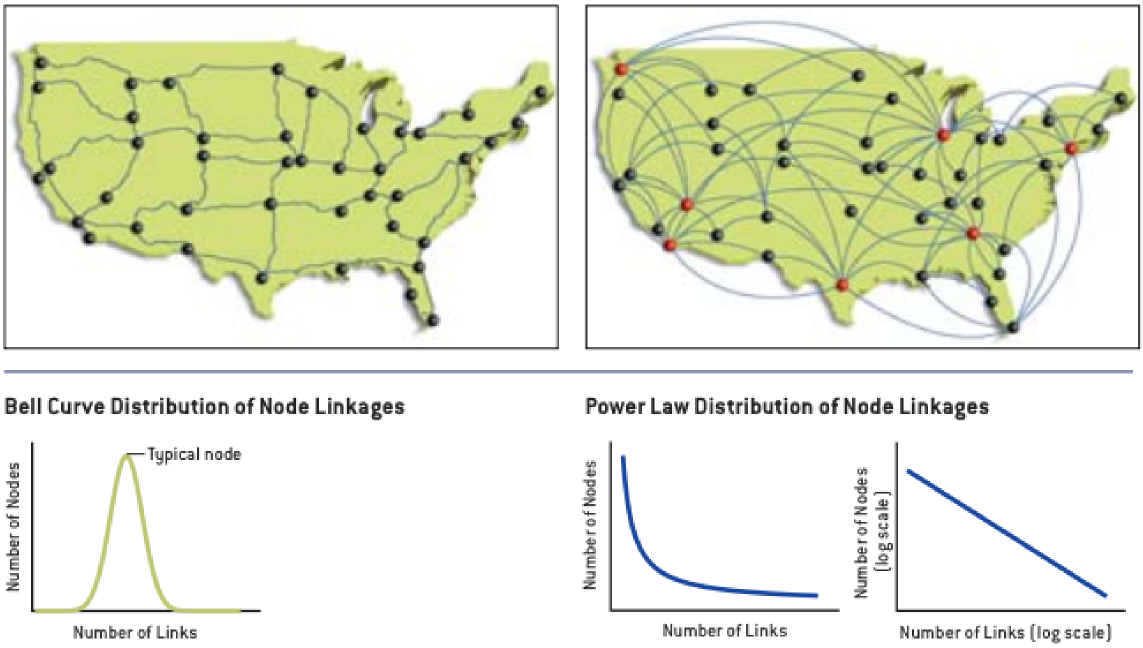
- non abbiamo ancora un modello con la distribuzione dei gradi "giusta"
- si può ottenere aggiungendo vertici secondo un modello di attaccamento preferenziale (preferential attachment)
- esempi
  - se ho molti amici, ho più possibilità di fare nuove amicizie
  - se ho un grosso giro d'affari mi è più facile farne di nuovi
  - molti preferiscono andare a cena in un ristorante molto frequentato
- sono fenomeni che amplificano le disuguaglianze

modello di Barabàsi-Albert

- si parte da due vertici connessi da uno spigolo
- a ogni passo si aggiunge un vertice v e lo si connette a uno dei precedenti
- la probabilità che un vecchio vertice u venga connesso a v è proporzionale al grado attuale di u
  - più precisamente, la probailità che u venga connesso a $v =$ (grado di u)/(somma dei gradi)
- i vertici che hanno grado alto lo incrementeranno sempre di più
- genera una distribuzione a legge di potenza
- alcuni vertici avranno molte connessioni (hub)



strade vs aerei



- distanza media piccola ok
- heavy tail ok
- clustering alto NO

bibliografia

- L. Barabási, Link, Einaudi 2004
- D. Watts, Six Degrees, Norton 2004