



SAPIENZA  
UNIVERSITÀ DI ROMA

DIPARTIMENTO DI INGEGNERIA INFORMATICA AUTOMATICA E GESTIONALE  
ANTONIO RUBERTI

# Sistemi Informativi Aziendali

*Umberto Nanni*

## Introduzione alle Reti Sociali

# Oltre 50 anni, portati bene

## ON A CLASS OF SKEW DISTRIBUTION FUNCTIONS

*Carnegie Institute of Technology*

1955

### I. INTRODUCTION

It is the purpose of this paper to analyse a class of distribution functions that appears in a wide range of empirical data—particularly data describing sociological, biological and economic phenomena. Its appearance is so frequent, and the phenomena in which it appears

Nessuno suppone che ci sia una connessione tra i calci di cavallo subiti da soldati dell'esercito tedesco e le cellule del sangue sulla lastra di un microscopio, a parte il fatto che lo stesso schema di casualità fornisce un modello astratto soddisfacente per entrambi i fenomeni.

so diverse, that one is led to the conjecture that if these phenomena have any property in common it can only be a similarity in the structure of the underlying probability mechanisms. The empirical distributions to which we refer are: (A) distributions of words in prose samples by their frequency of occurrence, (B) distributions of scientists by number of papers published, (C) distributions of cities by population, (D) distributions of books suffered by soldiers in the German army and blood cells on a microscope slide other than that the same urn scheme provides a satisfactory abstract model of both phenomena. It is in the same direction that we shall look for an explanation of the observed close similarities among the five classes of distributions listed above.

The observed distributions have the following characteristics in common:

(a) They are J-shaped, or at least highly skewed, with very long upper tails. The tails can generally be approximated closely by a function of the form

$$f(i) = (a/i^k) b^i, \quad (1.1)$$

# Una citazione

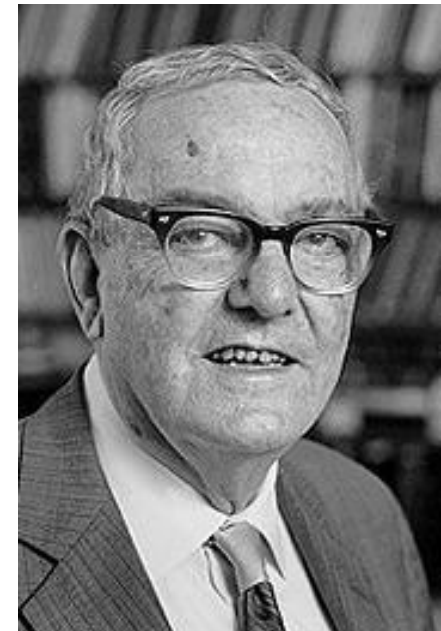
*“Se tracciamo una mappa di una interazione sociale – es., chi parla con chi – gli agglomerati di interazione densa sulla mappa identificheranno una struttura gerarchica alquanto ben definita.”*

*The Sciences of the Artificial*

Cambridge, MA, MIT Press, 1969.

**Herbert Alexander Simon [1916 – 2001]**

- 1975: premio Turing
- 1978: premio Nobel in Economia



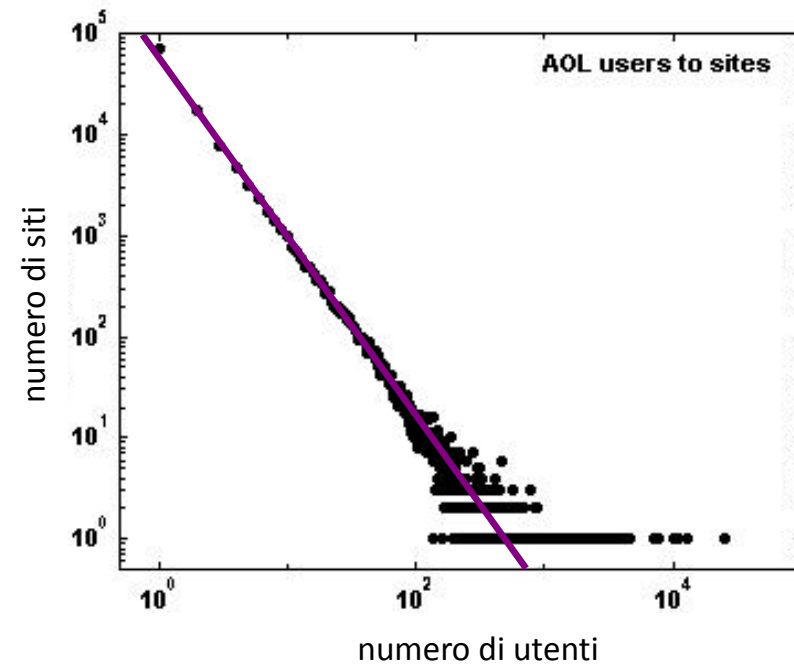
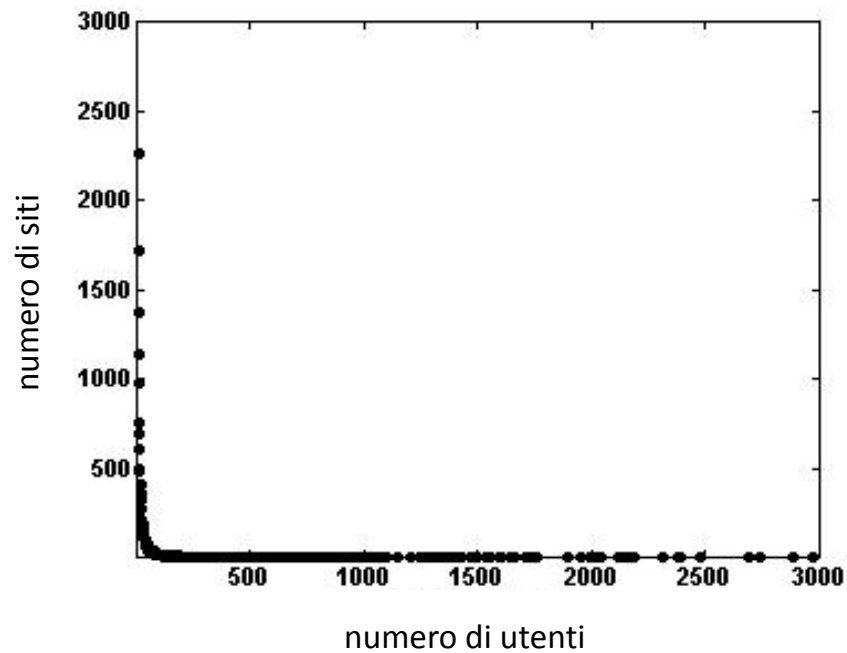
# Power Law (legge di Pareto, legge di Zipf)

- popolazione delle città
- magnitudo dei terremoti
- crateri della luna
- macchie solari
- dimensione dei file
- “dimensione” delle guerre
- frequenza dell'uso delle parole
- frequenza dei nomi propri
- numero di articoli scritti dagli scienziati
- numero di citazioni di articoli
- reddito pro-capite
- numero di specie nelle tassonomie
- accessi a pagine sul web
- vendite di libri, pezzi musicali, molti dei prodotti in vendita sul web
- ...

$$y = c x^{-\alpha} + \varepsilon(\dots)$$

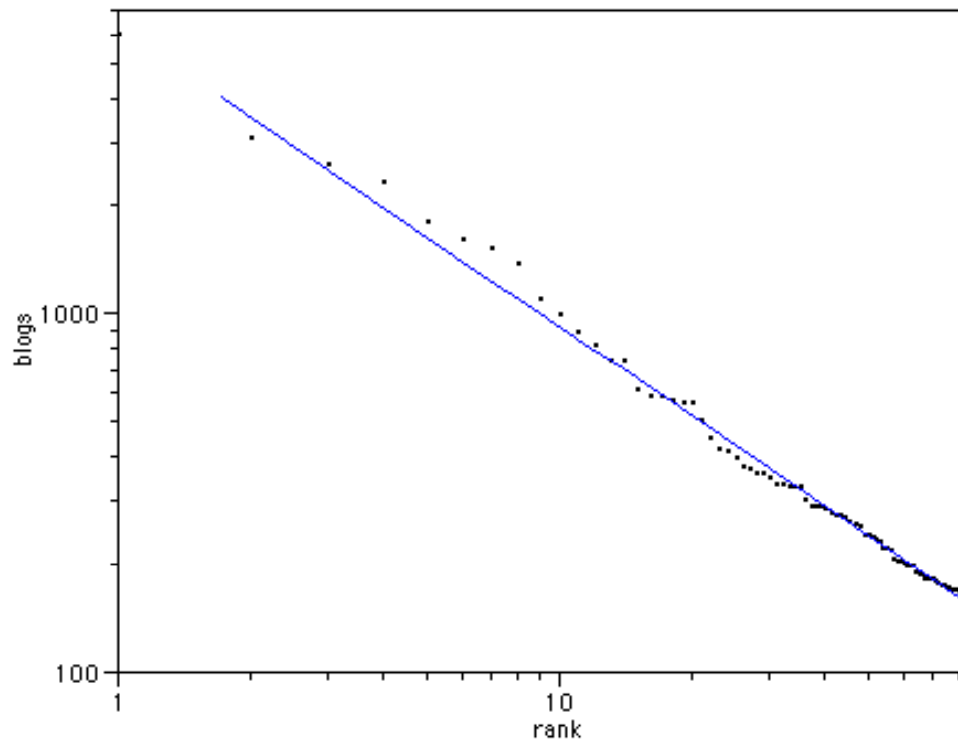
# Alcuni esempi

Una giornata di accessi al web da parte di utenti di AOL

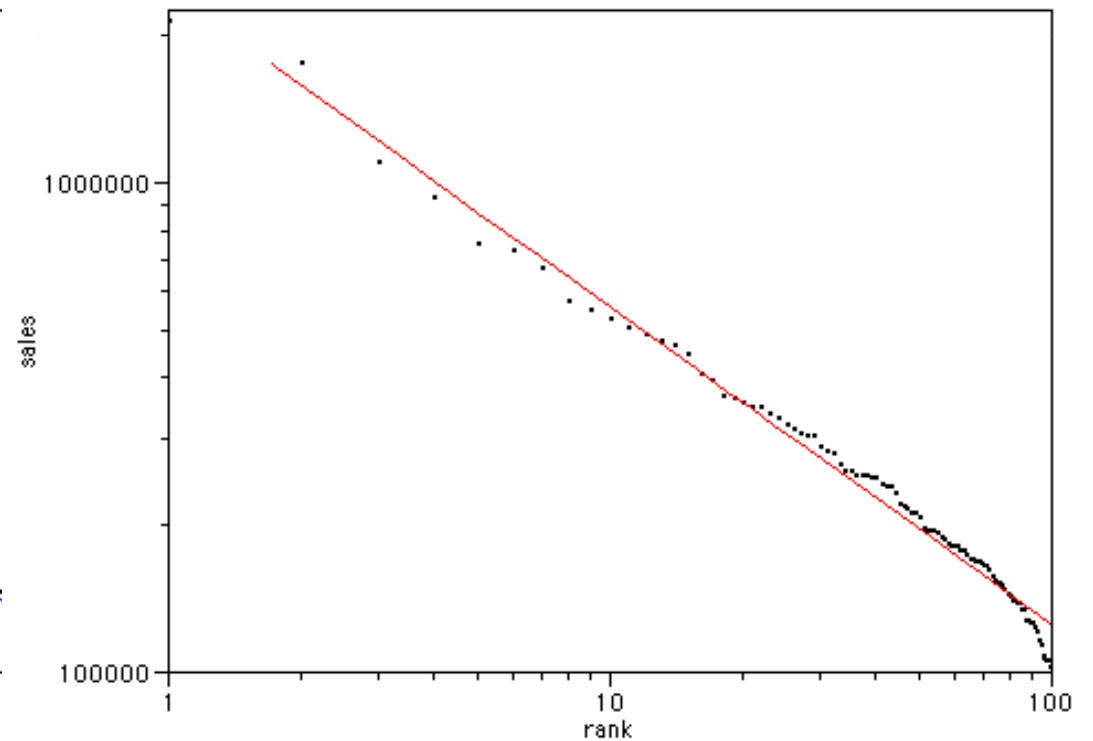


# Alcuni esempi (log-log)

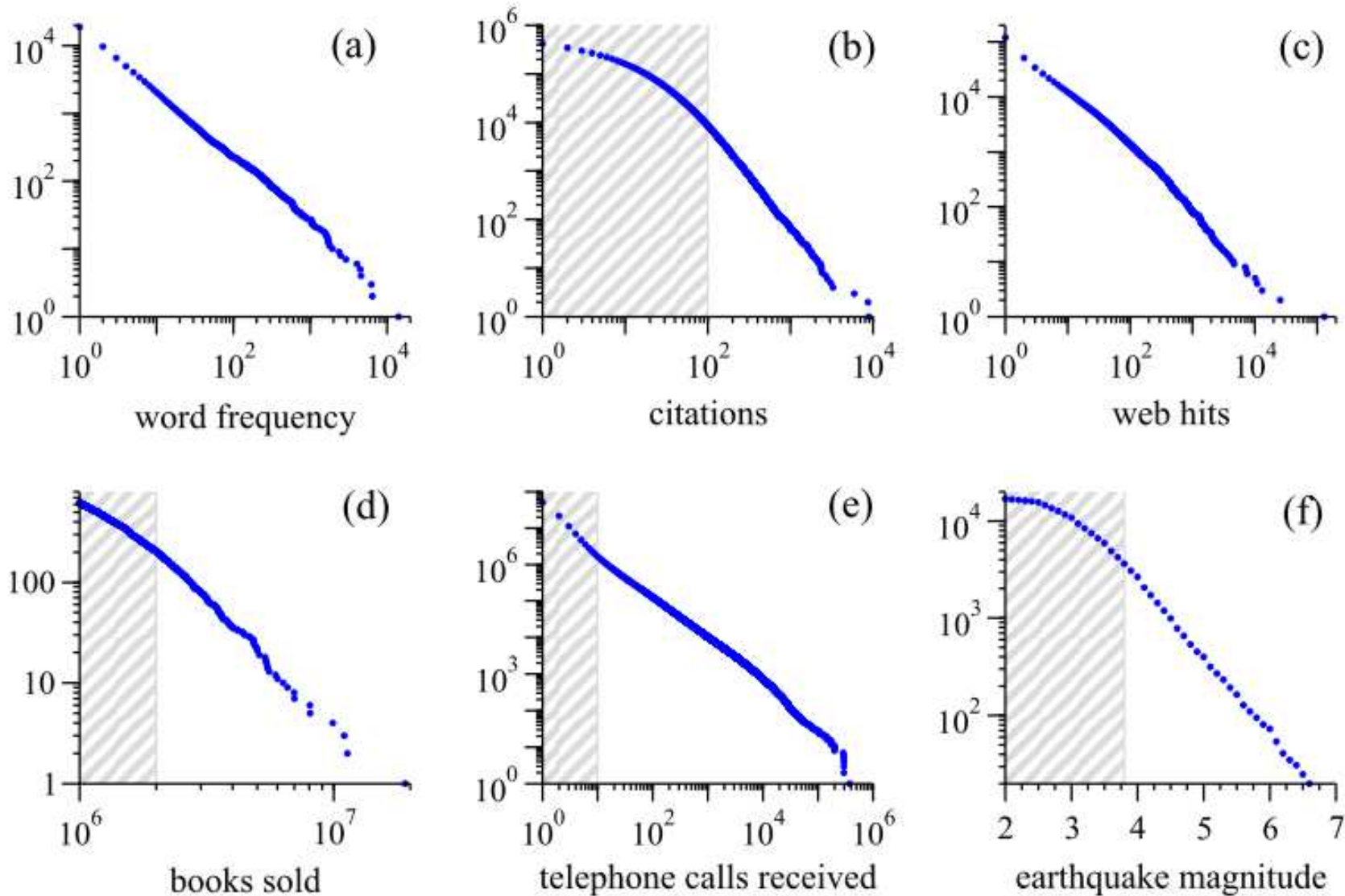
Incoming Blog links by rank



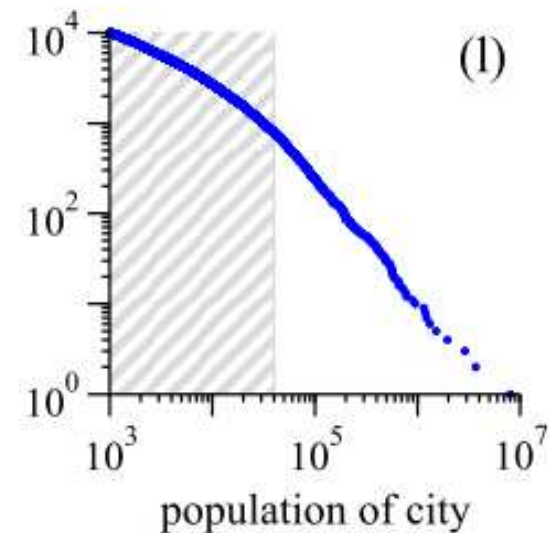
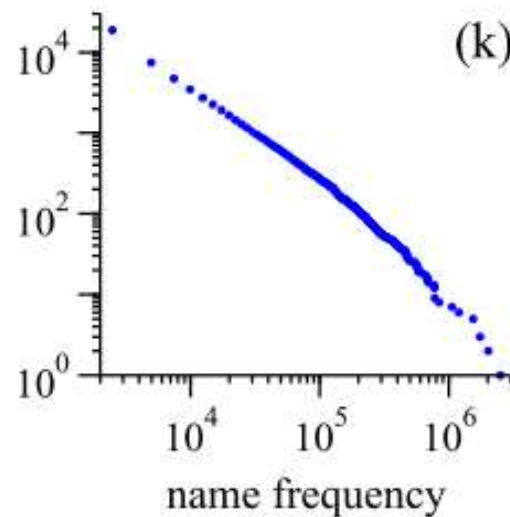
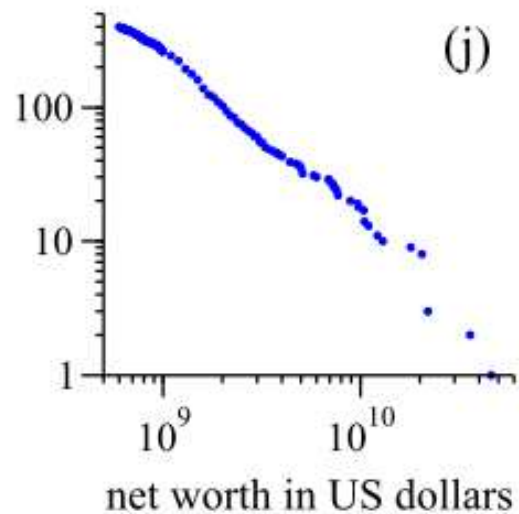
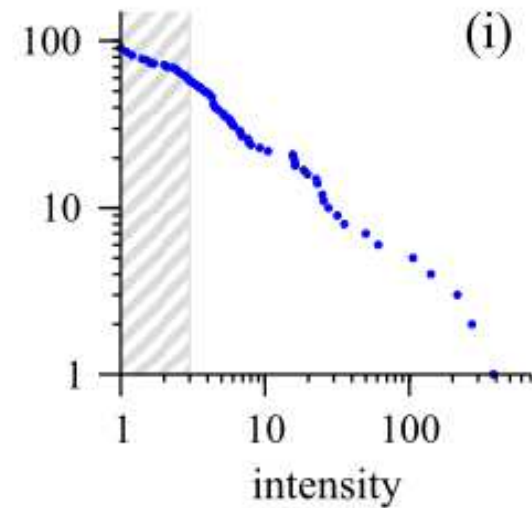
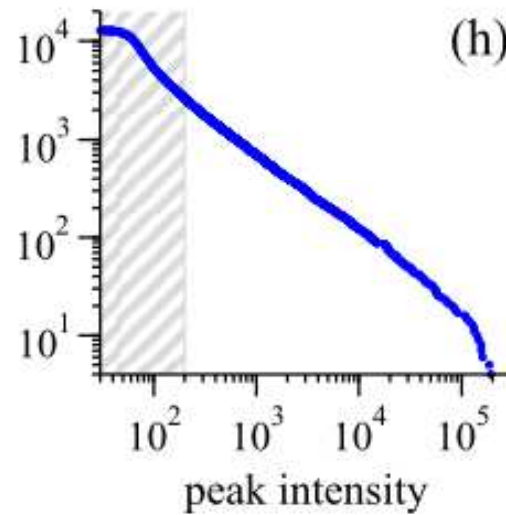
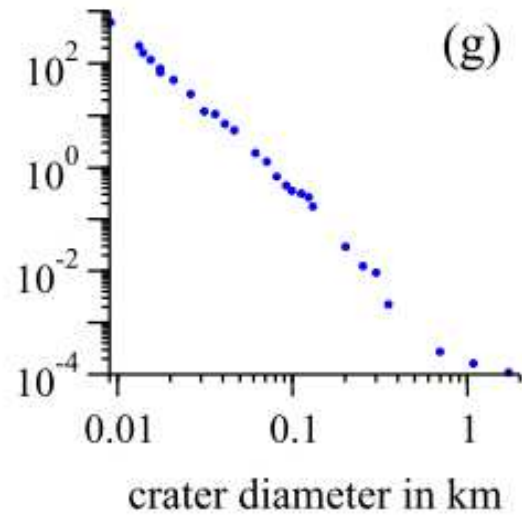
Vendite dei maggiori 100 giornali in USA



# Alcuni esempi



# Alcuni esempi





# Parametri nella power law

quantity	minimum $x_{\min}$	exponent $\alpha$
(a) frequency of use of words	1	2.20(1)
(b) number of citations to papers	100	3.04(2)
(c) number of hits on web sites	1	2.40(1)
(d) copies of books sold in the US	2 000 000	3.51(16)
(e) telephone calls received	10	2.22(1)
(f) magnitude of earthquakes	3.8	3.04(4)
(g) diameter of moon craters	0.01	3.14(5)
(h) intensity of solar flares	200	1.83(2)
(i) intensity of wars	3	1.80(9)
(j) net worth of Americans	\$600m	2.09(4)
(k) frequency of family names	10 000	1.94(1)
(l) population of US cities	40 000	2.30(5)

# Alcuni opportunisti

- base di clienti consistente
- conoscenza specifica di: una esigenza e del momento in cui questa si manifesta

up-seller:

**amazon.com**

	\$ millions	% change from 2005
Revenues	10,711.0	26.2
Profits	190.0	-47.1

cross-seller:

**Google™**

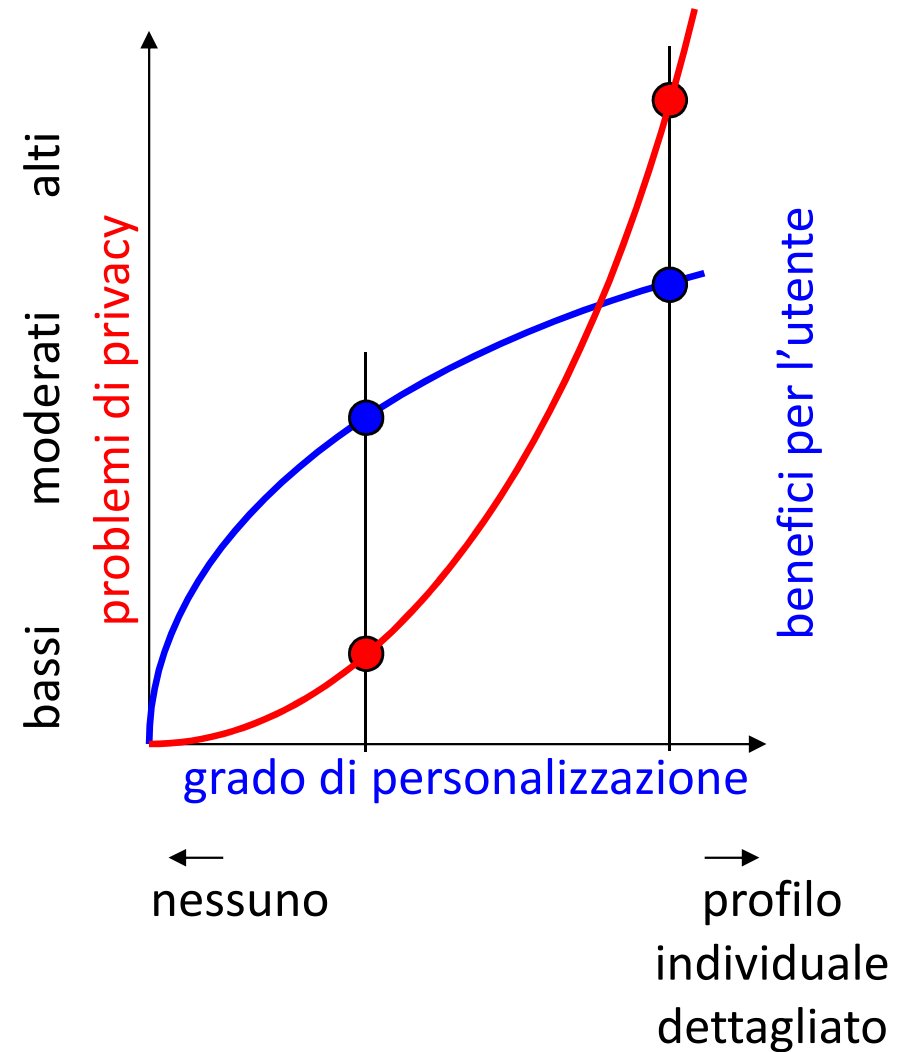
	\$ millions	% change from 2005
Revenues	10,604.9	72.8
Profits	3,077.4	110.0

# Personalizzazione e privacy [Barry Smyth]

## Esigenze contrastanti

problemi di privacy

benefici per l'utente  
(o cliente, o cittadino)



# Alcune osservazioni

- una base di (centinaia di) milioni di clienti/utenti (o visitatori sul sito) è una risorsa di elevato potenziale, specie se possiamo tracciare transazioni (acquisti, ricerche, etc.) o relazioni (“amici”, “mi piace”, etc.)
- l'appartenenza ad una social network può essere consapevole (es., con adesione esplicita) o inconsapevole (es., collocazione in gruppi di comportamento relativamente omogeneo, sulla base di informazioni transazionali)
- seppur con una certa differenziazione, esiste una diffusa pratica (con una accettazione più o meno consapevole) a “cedere” informazioni personali in cambio di servizi o vantaggi di varia tipologia

# Rete Sociale

---

NODI: insieme di soggetti:

persone, organizzazioni, animali, oggetti, ...

(in linea di massima: dati anagrafici)

ARCHI:

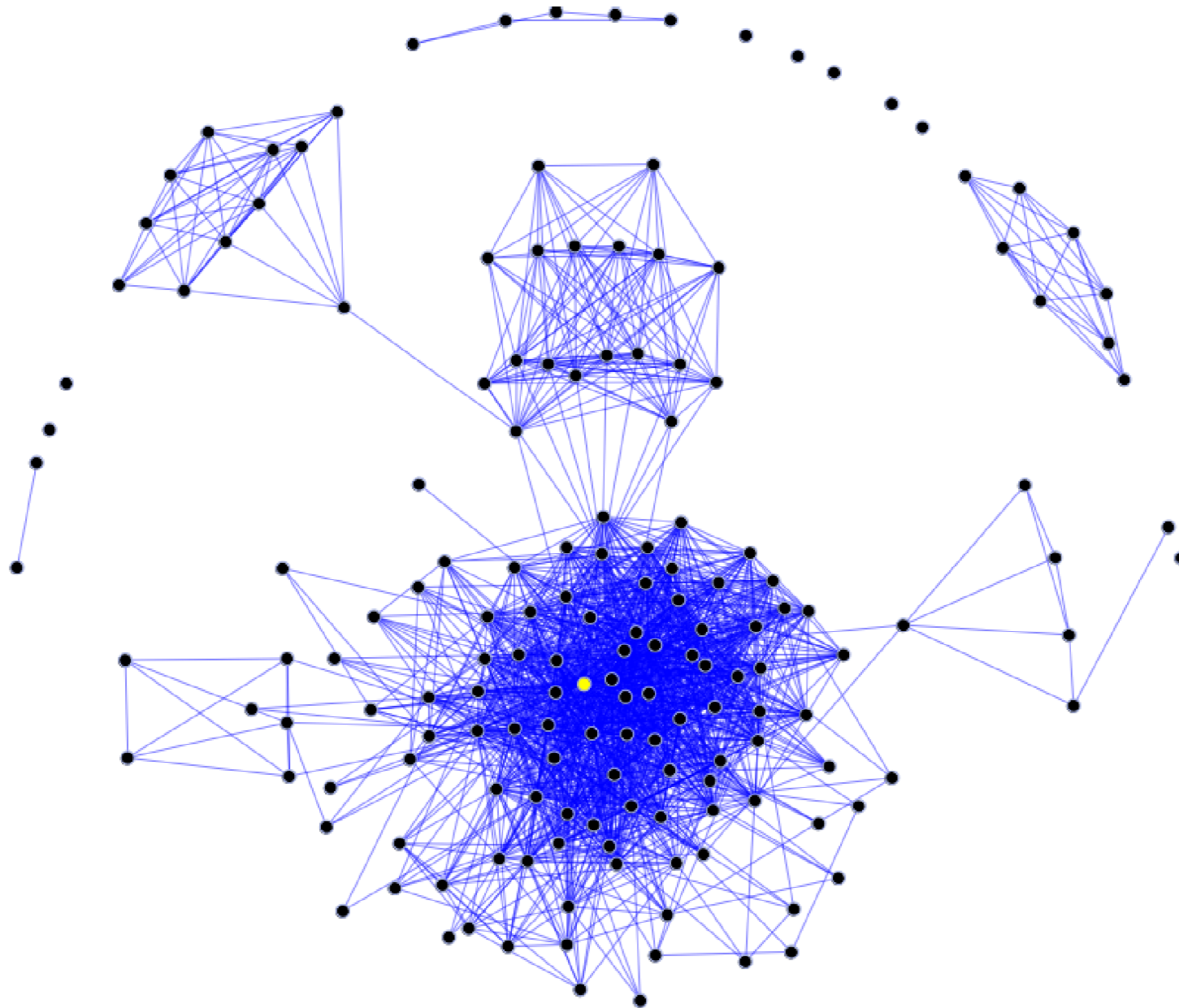
relazione binaria tra istanze facenti parte di un determinato insieme

(in linea di massima: dati transazionali)

# Esempi di Social Networks

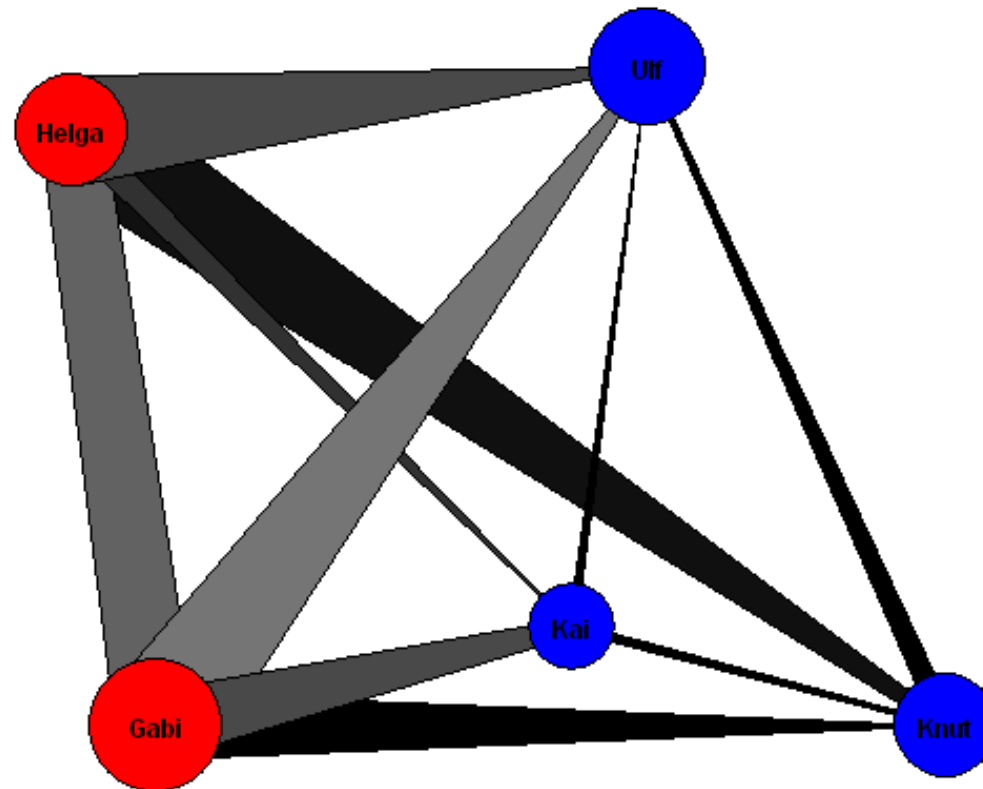
- World Wide Web
- Internet
- citazioni tra articoli scientifici
- citazioni tra autori
- conoscenza diretta tra individui
- relazioni di affari tra aziende
- invio di e-mail
- biologia: partecipazione di proteine ad un processo
- comunicazioni telefoniche
- ...

# Rapporti interpersonali



# Sociogramma

Rappresentazione grafica di un grafo sociale con metafora che fornisce una evidenza quantitativa del peso degli archi, unitamente al verso.





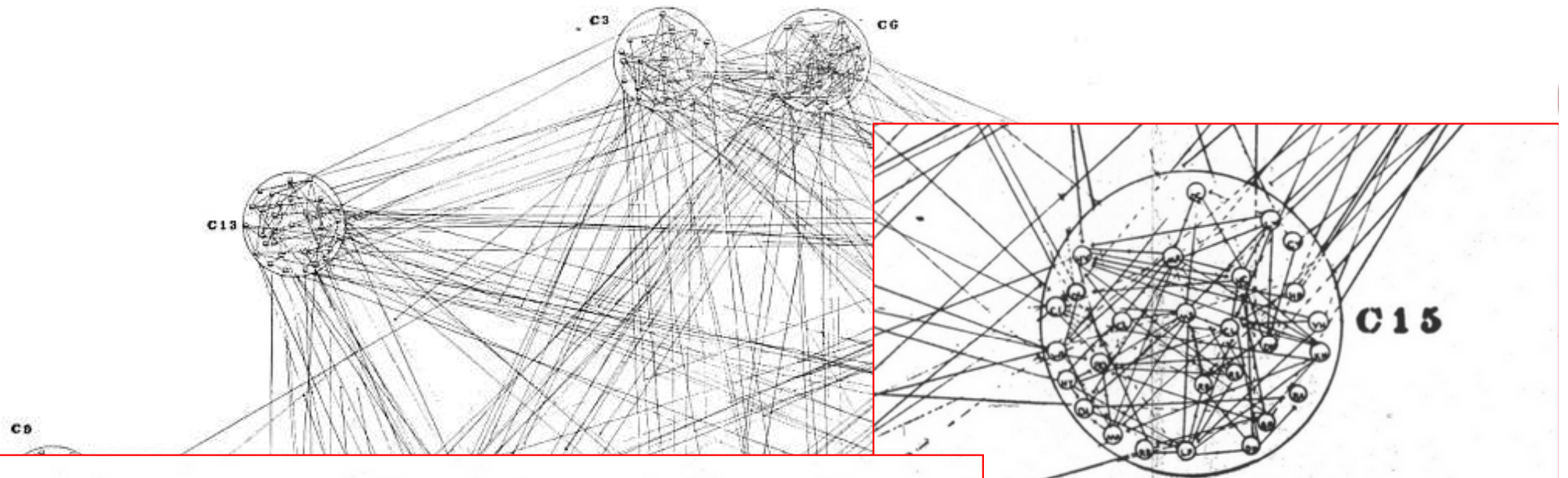
# Sociogramma

*“I sociometristi hanno ideato un processo di creazione di grafici, il sociogramma, che è più di un semplice metodo di presentazione. È prima di tutto un metodo di esplorazione. Rende possibile l'esplorazione di fatti sociometrici.*

*Il corretto posizionamento di ogni uomo e di tutte le interrelazioni di individui può essere visualizzato su un sociogramma. È attualmente l'unico metodo disponibile che rende l'analisi strutturale di una comunità possibile”.*

**Jacob Levy Moreno, 1934 (!)**

# SOCIOMETRIC GEOGRAPHY OF A COMMUNITY — MAP III



## LEGEND

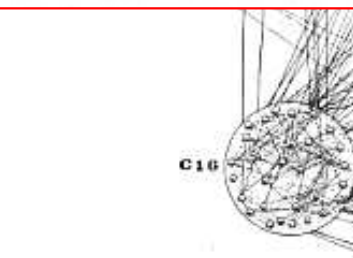
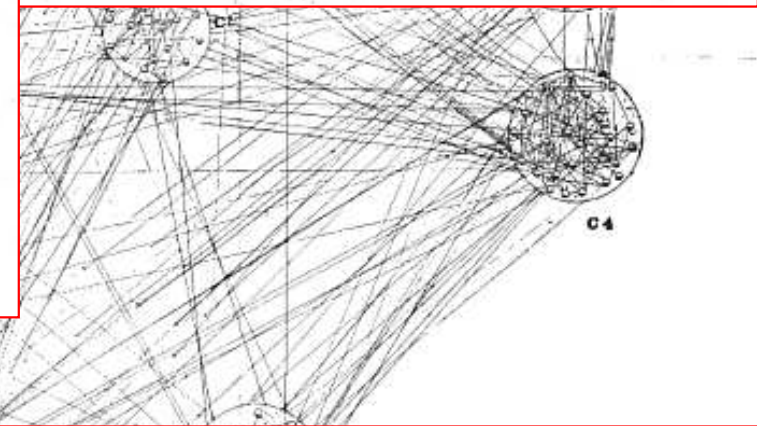
Population : 435 inhabitants.

Large circles : houses.

Small circles : individuals.

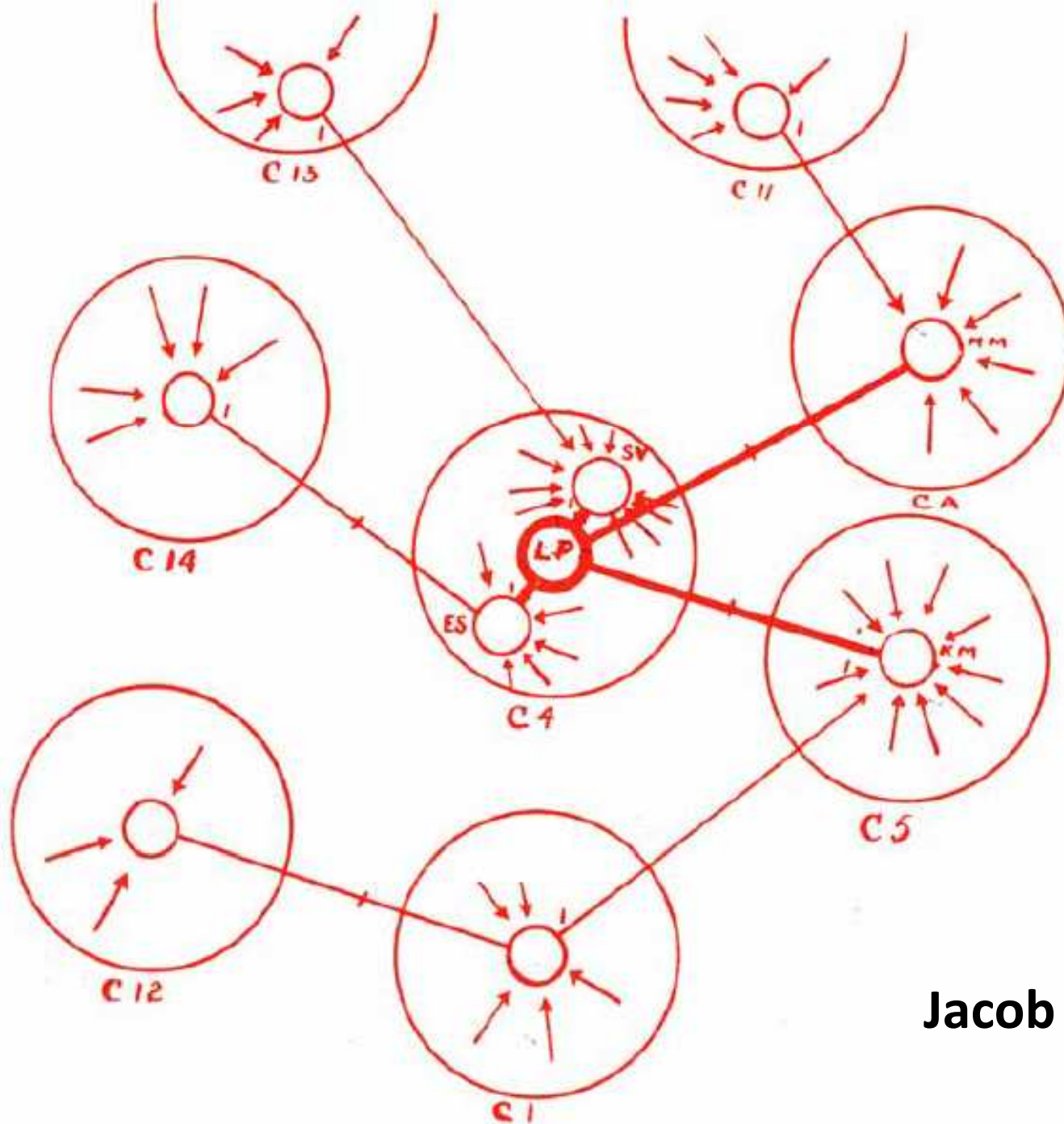
Each of the 4350 lines (tele) connects two individuals.

The chart portrays the houses in which they actually live and the houses in which they would like to live.



Population : 435 inhabitants.  
Large circles : houses.  
Small circles : individuals.  
Each of the 4350 lines (tele) connects two individuals.  
The chart portrays the houses in which they actually live and the houses in which they would like to live.

The original of this map is at the Moreno Institute, New York City. It was drafted in the fall of 1932 by O. and L. Osip, under the supervision of J. L. Moreno. The original is in three colors, red, blue and black and covers an area approximately ten times as large as this print.



**Jacob Levy Moreno, 1934**

Fig. 2. A Powerful Individual, LP. LP of C4 is the center of attraction from four individuals: SV and ES both of C4, MM of CA, and

# Esperimenti storici su una Social Network

## 1967: Esperimento di Milgram

- lettere inviate a destinatario specifico
- inviabili solo a destinatario personalmente conosciuto
- 5% successo; distanza-media=6

## 1969: Esperimento di Travers-Milgram

- rispetto al precedente: più informazioni sul destinatario
- 29% successo; distanza-media=5,2

**Fenomeno del “Mondo Piccolo”  
(Small World Phenomenon)**

# Esempio di Rete Sociale nelle TLC

<b>nodo:</b>	<b>soggetto di interesse</b>
<b>arco:</b>	<b>frequenza / volume di comunicazione in una determinata finestra temporale</b>
<b>arco pesato:</b>	<b>valore di misura</b>
<b>arco non pesato:</b>	<b>superamento di soglia</b>
<b>problemi:</b>	<b>soglie quantitative</b> <b>dinamica nel tempo</b>
<b>obiettivo:</b>	<b>identificazione di “comunità” e dei ruoli all’interno della comunità</b>



# Social Network – metriche e proprietà

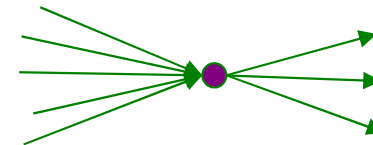
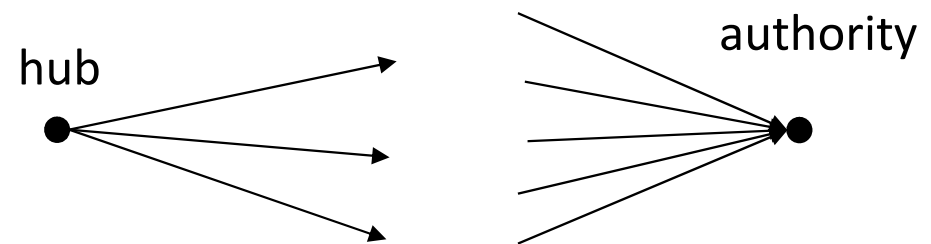
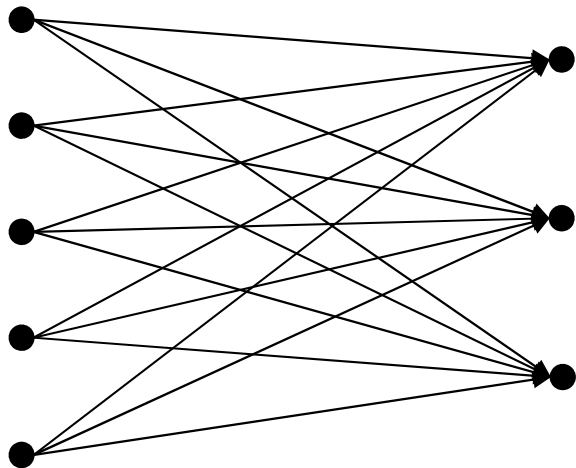
- **grado in/out  $g_{in}$ ,  $g_{out}$** : numero di archi entranti/uscenti dai nodi
- **diametro  $d$** : distanza massima tra una coppia di nodi → Small World Phenomenon
- **funzione di vicinato  $N(h)$** : numero di coppie di nodi entro raggio  $h$

# Comunità su un grafo sociale

Componenti bipartite:

- HUB: nodi con molti archi uscenti
- AUTHORITIES: nodi con molti archi entranti

componente bipartita completa



Social Networking Potential (SNP)

- caso specifico: utente alpha

# Altre metriche su Social Networks (1)

- **Betweenness** (“Immezzezza”): misura quanto un individuo si trova tra altri individui nella rete; la misura in cui un nodo è collegato direttamente solo a quelle di altri nodi che non sono direttamente collegati fra loro; un intermediario; liaisons; ponti.
- **Centralità:**
  - **Grado:** il numero di archi incidenti.
  - **Betweenness (“immezzezza”)** : il grado in cui un nodo appartiene ad un percorso minimo tra le coppie di nodi.
  - **dell’autovettore:** una misura dell'importanza di un nodo in una rete. Si assegna punteggi relativi a tutti i nodi della rete sulla base del principio che le connessioni ai nodi che hanno un punteggio elevato contribuiscono di più al punteggio del nodo in questione.
  - **Closeness (vicinanza):** (inverso della) somma delle distanze da tutti gli altri nodi nella rete
- **Centralizzazione:** differenza tra il grado di un nodo divisa per massimo possibile somma delle differenze. Una rete centralizzata avrà molti dei suoi collegamenti concentrati intorno uno o pochi nodi, mentre una rete decentrata è quello in cui c'è una piccola variazione tra il numero di collegamenti che ogni nodo possiede
- **Coefficiente di Clustering:** una misura della probabilità che due nodi adiacenti ad un nodo assegnato sono collegati tra loro. Un coefficiente di clustering più elevato indica una maggiore propensione all’aggregazione.



# Altre metriche su Social Networks (2)

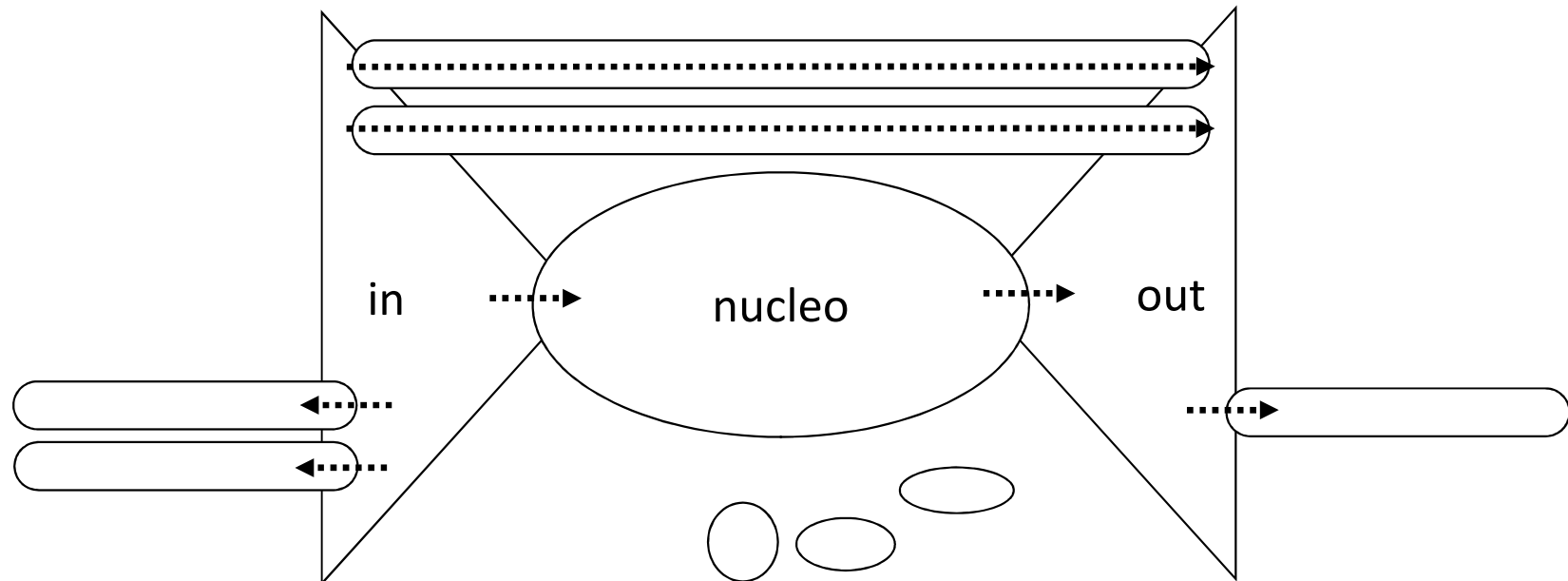
- **Coesione:** Il grado in cui gli attori sono collegati direttamente tra loro da legami coesivi. I gruppi sono identificati come 'cricche' (cliques) se ogni attore è direttamente legato ad ogni altro attore, 'circoli sociali' (comunità) in caso di rilassamento
- **Densità (globale) di una rete:** è il rapporto tra il numero di archi in una rete rispetto al numero massimo possibile (rete completa)
- **Densità locale di una comunità:** è il rapporto tra il numero di archi tra nodi della comunità rispetto al numero massimo possibile (rete completa)
- **Densità individuale (di un nodo in una comunità):** è il rapporto tra il numero di archi incidenti su un nodo (che lo connettono a nodi della comunità) rispetto al numero massimo possibile (dimensione della comunità)
- **Distanza tra due nodi:** lunghezza del percorso minimo tra i due nodi. Distanza media è la media di queste distanze tra tutte le coppie di nodi
- **Radialità:** misura con la quale un individuo è connesso agli altri e fornisce nuovi dati e di influenza
- **Reach (area raggiungibile):** il numero (o la proporzione) dei nodi raggiungibili
- **Coesione strutturale di una comunità:** Il numero minimo di nodi che, se rimosso da un gruppo, lo sconnette
- **Equivalenza strutturale (di due nodi):** misura in cui due nodi hanno un insieme di collegamenti con altri nodi della rete
- **Buco strutturale:** uno o più collegamenti non presenti che collegherebbero tra loro zone altrimenti non connesse (o ne diminuirebbero drasticamente la distanza).

# Alcune Social Network di interesse

- $G_{\text{int}}$  Grafo di internet
  - il grado segue la Power Law
- $G_{\text{web}}$  Grafo del web
  - il grado segue la Power Law
  - il diametro è circa 20 (Small World Phenomenon)
- $G_{\text{mail}}$  Grafo delle e-mail
  - il grado segue la Power Law

# Grafo del web

- “importanza” delle pagine → concetto sfruttato da Google (Page Rank)
- nuclei bipartiti (hub/authority)
- struttura generale: farfalla (bow-tie)

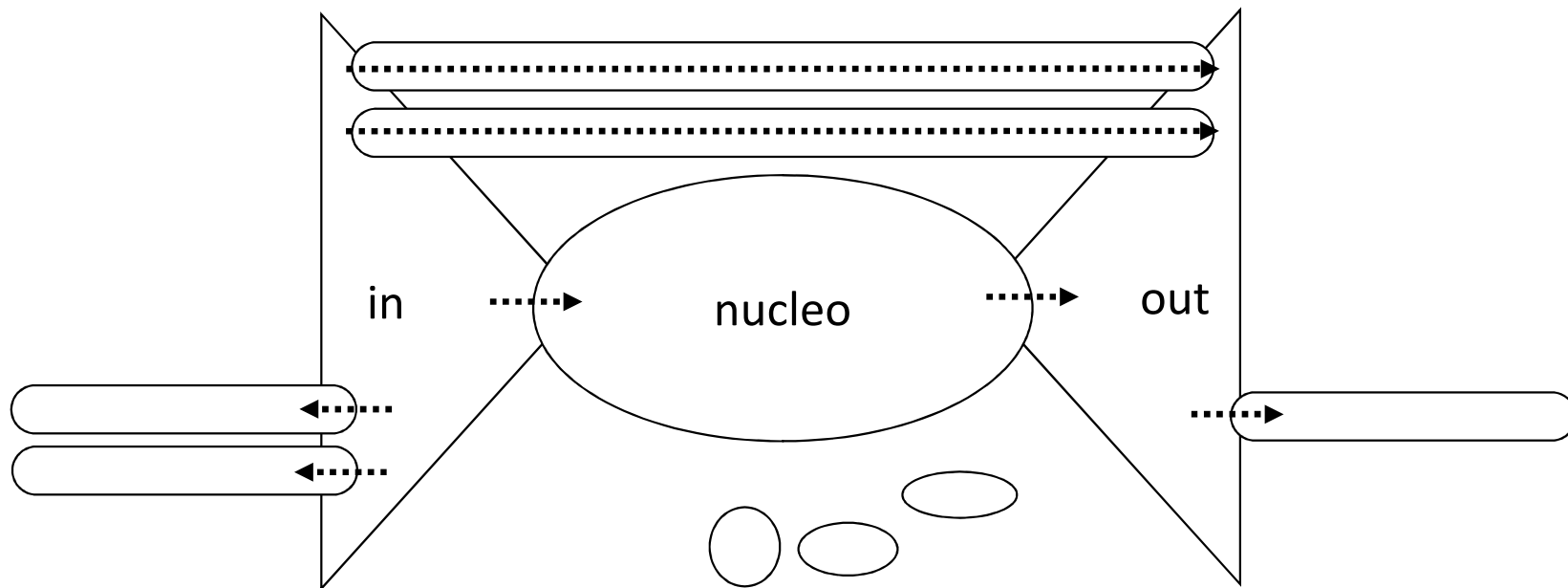


# Social Network su reti telefoniche

- telefonia fissa: proprietà poco focalizzate, perché il telefono è usato da diverse persone
- traffico mobile: proprietà più nette: si presta molto meglio all'analisi
- $G_{\text{call}}$ : Grafo delle chiamate
- $G_{\text{sms}}$ : Grafo degli SMS
- $G_{\text{sms}}$  è più “sociale” di  $G_{\text{call}}$

# Analisi del traffico telefonico (Soc. India, 53 mln ut.)

- Grafo delle chiamate:  $G_{\text{call}}$
- Grafo degli SMS:  $G_{\text{sms}}$



# Grafo delle Chiamate: $G_{\text{call}}$

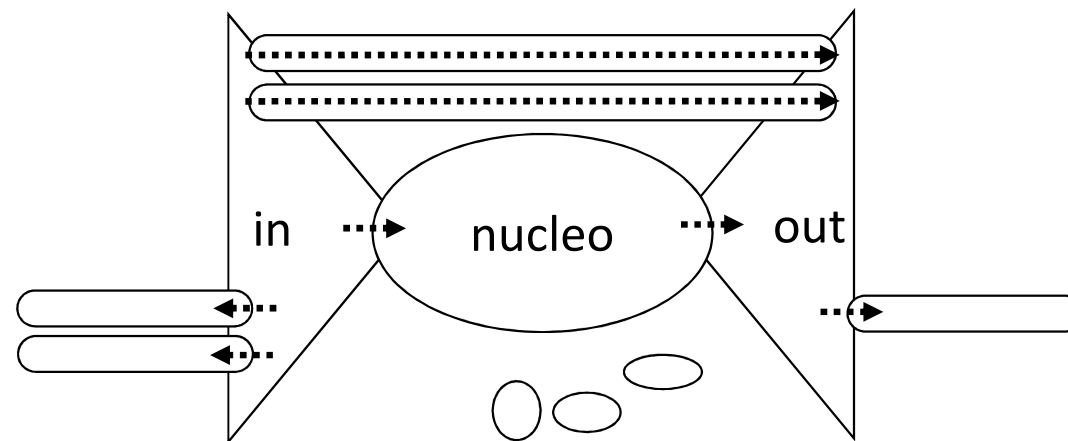
- la distribuzione del grado  $g_{\text{in}}, g_{\text{out}}$ : segue la Power Law
- $g_{\text{in}}$  è simile a  $G_{\text{int}}$ , ma  $g_{\text{out}}$  è diverso da  $G_{\text{int}}$  e  $G_{\text{mail}}$
- diametro  $< 13 \rightarrow$  Small World Phenomenon
- page-rank:  $p(i) = q/N + (1 - q) \sum_{[j:j \rightarrow i]} p(j)/d_{\text{out}}(j)$
- CFC: 1 gigante, poi segue la Power Law

# Alcune risultanze dall'analisi di $G_{call}$

## Componenti del Papillon

## nodì % del totale

- IN 8.5
- nucleo (SCC) 63
- OUT 18.7
- TENDRIL, TUBE and DISC 9.8



# Grafo degli SMS: $G_{\text{sms}}$

- $G_{\text{sms}}$
- ha metà dei nodi e un terzo degli archi di  $G_{\text{call}}$
- reciprocità ( $G_{\text{call}}$ )= 30%
- reciprocità ( $G_{\text{sms}}$ )= 60%
- distribuzione delle cricche: molto differente

tra  $G_{\text{call}}$  e  $G_{\text{sms}}$

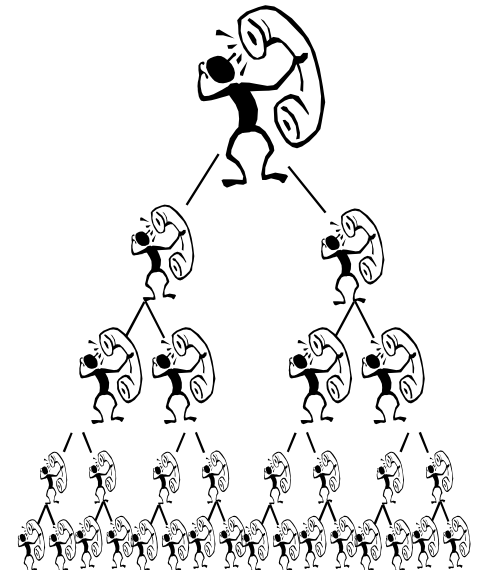


# Social Network – miniglossario

- **social marketing**: marketing di interesse sociale
- **word of mouth**: passaparola
- **influencer marketing**: basato su persone influenti
- **engagement marketing**: coinvolgimento di clienti nel marketing
- **viral marketing**: sfruttamento di una rete sociale per diffondere o promuovere un brand o prodotto
- **stealth marketing**: marketing fantasma (scorretto)
- **astroturfing**: forma di marketing fantasma che simula la manifestazione spontanea di una opinione positiva

# Principi di Marketing Virale

- 1. regala prodotti e/o servizi (non quelli oggetto del business)**
- 2. esiste un modo facile di comunicare ad altri il messaggio**
- 3. adotta una soluzione scalabile**
- 4. sfrutta motivazioni e comportamenti comuni**
- 5. utilizza forme di comunicazione esistenti**
- 6. approfitta delle risorse altrui**



# Un capolavoro di Marketing Virale

Tohato (Giappone) lancia sul mercato due tipi di patatine

- Habanero (piccante come il fuoco dell'inferno)
- Satan Jorquia (piccante come un bazooka mortale)

