Tommaso Russo - Informatica IV - Programma svolto

Testo di lettura:

Mark Buchanan
"NEXUS - la rivoluzionaria teoria delle reti"
Arnoldo Mondadori collana "saggi"
settembre 2003, ISBN 88-04-51251-2, euro 18.

Riferimento globale:

Reka Albert1 and Albert-Laszlo Barabasi
Statistical Mechanics of Complex Networks

reperibile a: http://arxiv.org/abs/cond-mat/0106096
(verrà distribuito in formato elettronico)

21.4.2004

Introduzione al problema delle reti "Piccolo Mondo"

L' oracolo di Kevin Bacon: http://www.cs.virginia.edu/oracle/
"Posizione privilegiata" di Kevin Bacon?
"Posizione privilegiata" di qualsiasi altro attore: http://www.cs.virginia.edu/oracle/star_links.html

Il (Mito? Leggenda metropolitana? Problema?) "six degrees of separation"

Stanley Milgram: gli esperimenti (ca 1960) (vedi NEXUS, pg 21-28)

- "Obbedienza al carnefice" (Vedi <lettura non obbligatoria> "The Perils of Obedience", 1974 by Stanley Milgram. Distribuito.)
- "la lettera persa"
- "lettere nel piccolo mondo"

Obiezioni all' esperimento di Milgram (percentuale lettere arrivate), ripetizione dell' esperimento in altri contesti.

Problema: quale modello può spiegare la rete "piccolo mondo"?

Prime ipotesi da parte degli studenti: propagazione delle reti di conoscenza ad albero:
- binario (non porta a 6 ma a **34** gradi di separazione)
- Ennario - N = ? Nmin = Radice sesta (6.000.000.000) = ca 43
  - cosi' pero' esiste UN SOLO path da me a Kevin Bacon
  - e come si fa a trovarlo? (Problema dell' algoritmo di ricerca)
- Valori ragionevoli di conoscenti per essere umano? (50 - 100 - 200...)

Primo approccio: reti casuali

Primi richiami di teoria dei grafi (da completare in seguito)

Teoremi strani: “proprietà vera per **quasi tutti i** grafi che soddisfano a...” - definizione esatta:

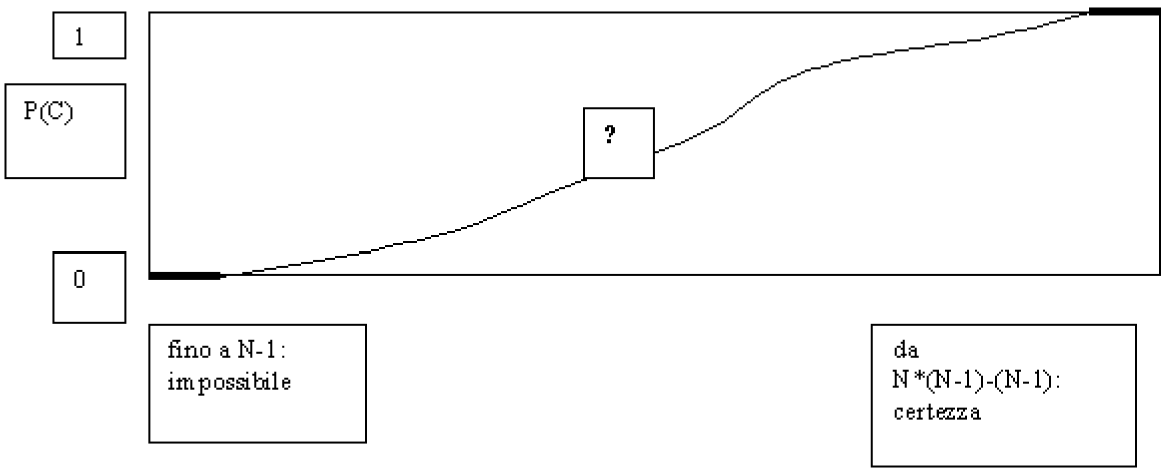
One says that ***almost every graph*** *has property Q* (in a certain model) if the probability that a graph in the model satisfies *Q* goes to 1 when *n* ?   ? (of course, in this case *p* and *M* might depend on *n*).

Matematica “sperimentale”: il limite p -> 1 per N -> infinito ci dice poco; è importante che 1-p sia trascurabile agli effetti pratici per reti di dimensione pratica. Evidenza computazionale: se già per N “piccolo” 1-p è trascurabile, allora per N grandi si può aspettarsi p = ca 1 “quasi certezza”.

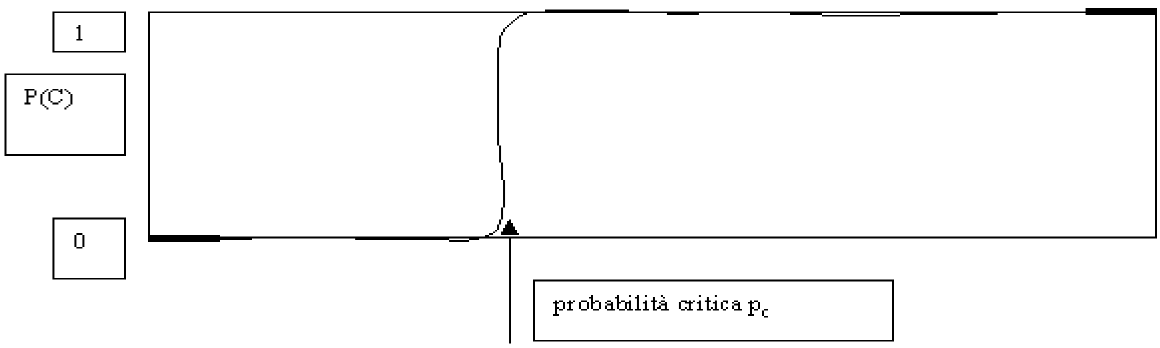
Grafo casuale di probabilità p: esistono p * N * (N-1) /2 archi fra gli N * (N-1) /2 possibili. Quand’ e’ che il grafo è connesso?

Teorema fondamentale della connessione di Erdos - Reny:

Cosa sappiamo:

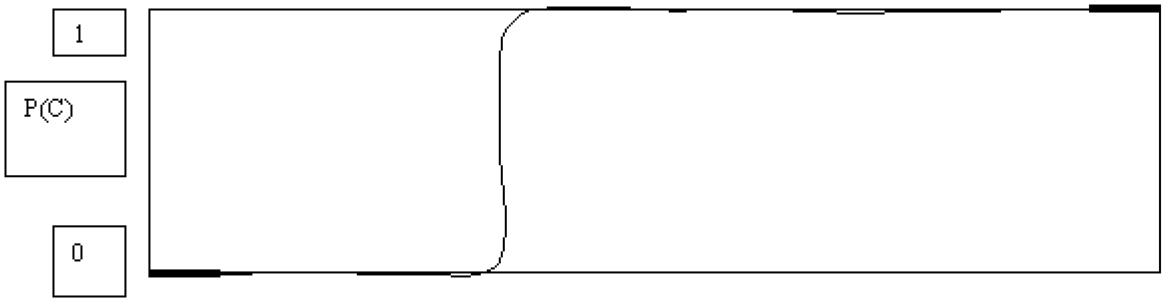


Cosa ha dimostrato Erdos:



22.4.2004

in dettaglio: la proprietà C è qui l' essere connesso, ma lo stesso andamento con diversa p_c si ha anche per altre proprietà (p.es. contenere deteminati subgrafi):



p: 0 p_c 1

n: 0 nc = p_c N (N-1)/2 N (N-1)/2

1. punto di flesso a p_c = (ln N) / N, ossia n_c = ln N (N-1)/2; grado totale grafo = K_c = 2 n_c ; grado medio dei nodi $<k_c>$ = K_c / N = ca ln N = p_c N

2. flesso praticamente verticale

3. P(C) trascurabile per p < p_c ; 1 - P(C) trascurabile per p > p_c

Dalle formule reperibili su
Reka Albert1 and Albert-Laszlo Barabasi Statistical Mechanics of Complex Networks
pg. 13 (dimostrate nelle pagg. precedenti, dimostrazioni NON da sapere)

si ricavano le seguenti tabelle:

Probabilità critica perche' un grafo casuale risulti connesso					
N	pcrit=ln(N)/N	n max= N(N-1)/2	n = pcrit * n max	<k>= 2n/N =pcrit*N	d(p crit) = lnN/ln<k>
5	3,22E-001	10	3,218875825	1,61	3,38
10	2,30E-001	45	10,36163292	2,30	2,76
15	1,81E-001	105	18,95635141	2,71	2,72
20	1,50E-001	190	28,4594566	3,00	2,73
25	1,29E-001	300	38,6265099	3,22	2,75
50	7,82E-002	1225	95,84456363	3,91	2,87
100	4,61E-002	4950	227,9559242	4,61	3,02
1.000	6,91E-003	499500	3450,423762	6,91	3,57
10.000	9,21E-004	49995000	46047,09669	9,21	4,15
100.000	1,15E-004	4999950000	575640,5168	11,51	4,71
1.000.000	1,38E-005	5E+11	6907748,371	13,82	5,26
1.000.000.000	2,07E-008	5E+17	10361632908	20,72	6,84
6.000.000.000	3,75E-009	1,8E+19	67545075907	**22,52**	**7,23**
1.000.000.000.000	2,76E-011	5E+23	1,38155E+13	27,63	8,33

Probabilità critica perche' un grafo casuale contenga un ciclo e un componente gigante					
N	Pcrit=1/N	n max= N(N-1)/2	n = p crit * n max	<k>= 2n/N =pcrit*N	N del comp. Gig.
5	2,00E-001	10	2	1,00	2,92
10	1,00E-001	45	4,5	1,00	4,64
15	6,67E-002	105	7	1,00	6,08
20	5,00E-002	190	9,5	1,00	7,37
25	4,00E-002	300	12	1,00	8,55
50	2,00E-002	1225	24,5	1,00	13,57
100	1,00E-002	4950	49,5	1,00	21,54
1.000	1,00E-003	499500	499,5	1,00	100,00
10.000	1,00E-004	49995000	4999,5	1,00	464,16
100.000	1,00E-005	4999950000	49999,5	1,00	2154,43
1.000.000	1,00E-006	5E+11	499999,5	1,00	10000,00
1.000.000.000	1,00E-009	5E+17	499999999,5	1,00	1000000,00
6.000.000.000	1,67E-010	1,8E+19	3000000000	1,00	3301927,25
1.000.000.000.000	1,00E-012	5E+23	5E+11	1,00	100000000,00
Probabilità critica perche' un grafo casuale contenga un albero di ordine 3					
N	Pcrit=N **(-3/2)	n max=N(N-1)/2	n = pcrit * n max	<k>= 2n/N =pcrit*N	
5	8,94E-002	10	0,894427191	0,45	
10	3,16E-002	45	1,423024947	0,32	
15	1,72E-002	105	1,807392228	0,26	
20	1,12E-002	190	2,124264579	0,22	
25	8,00E-003	300	2,4	0,20	
50	2,83E-003	1225	3,464823228	0,14	
100	1,00E-003	4950	4,95	0,10	
1.000	3,16E-005	499500	15,79557691	0,03	
10.000	1,00E-006	49995000	49,995	0,01	
100.000	3,16E-008	4999950000	158,1123019	0,00	
1.000.000	1,00E-009	5E+11	499,9995	0,00	
1.000.000.000	3,16E-014	5E+17	15811,38829	0,00	
6.000.000.000	2,15E-015	1,8E+19	38729,83346	0,00	
1.000.000.000.000	1,00E-018	5E+23	500000	0,00	

Evoluzione di un grafo casuale. Fasi principali:

1- connessioni sparse, nessun albero (perchè?)
2- connessioni sparse con qua e là alberi, nessun ciclo (perchè?)
3- ciclo e componente “gigante”; accrescimento della componente gigante e assorbimento
4- grafo connesso

Importanza della componente gigante nella teoria della percolazione (cenno)
rif Reka Albert1 and Albert-Laszlo Barabasi Statistical Mechanics of Complex Networks
pg. 14 e succ., da darci un’ occhiata

I grafi casuali spiegano la struttura delle reti “piccolo mondo”?
Al contrario di quanto afferma Buchanan a pag. 27 in Nexus, sembrerebbe di si’, almeno per quanto riguarda i gradi di separazione; vedi il diametro (MAX fra i gradi di separazione calcolati) nella prima tabella.

Si consideri inoltre che:

1- “6 gradi di separazione” è dato come MEDIA (non è un diametro) ed è una misura empirica limitata, relativa alle lettere ARRIVATE nell’ esperimento di Milgram; magari quelle che hanno seguito un path più lungo si sono perse;
2- nella rete delle relazioni sociali <k> è senz’ altro superiore a 23, il che corrisponde a una $p > p_c$ e quindi diametro inferiore;

Per quanto riguarda 1 - ,

I dati empirici possono essere ricavati da misure esatte di reti esistenti diverse da quella delle relazioni sociali umane: vedi lista :

Reka Albert1 and Albert-Laszlo Barabasi Statistical Mechanics of Complex Networks
pg. 8
Esaminate con attenzione le colonne relativi ai path medi, $\boldsymbol{l}$ e $\boldsymbol{l}_{real}$

Per quanto riguarda 2 - ...

28.4.2004

Probabilità critica perche' un grafo casuale risulti connesso				
N	<k>= 2n/N =pcrit*N	d(p crit) = lnN/ln<k>	d(2*pcrit)	d(3*pcrit)
5	1,609437912	3,381989195	1,376726788	1,022192
10	2,302585093	2,760785994	1,507736912	1,191417
15	2,708050201	2,718301206	1,602988372	1,292723
20	2,995732274	2,730371059	1,67327947	1,3643
25	3,218875825	2,753453576	1,72855306	1,41948
50	3,912023005	2,867937186	1,901623351	1,588531
100	4,605170186	3,015473824	2,074095657	1,753821
1.000	6,907755279	3,574249917	2,630732177	2,278842
10.000	9,210340372	4,148191314	3,161291439	2,775086
80.000	11,28978191	4,657696664	3,621949968	3,205039
1.000.000	13,81551056	5,261464354	4,162628527	3,709455
300.000.000	19,51929303	6,569048577	5,326517618	4,795877
400.000.000	19,80697511	6,633204473	5,38352783	4,849128
1.000.000.000	20,72326584	6,836525469	5,564182808	5,017899
6.000.000.000	22,51502531	7,229834018	5,913599547	5,344439
1.000.000.000.000	27,63102112	8,325257055	6,886945897	6,254827

Come si vede, aumentando la densità di connessione 2 - 3 volte (44 e 66 sono più vicini di 22 alla stima di quanti conoscenti abbia un essere umano) il diametro diminuisce, ma di poco; quello che aumenta molto è la resistenza della connettività alla distruzione di connessioni e anche di nodi (che distrugge le connessioni afferenti ai nodi e per di più diminuisce N, aumentando la probabilità critica per la connettività e quindi rendendo facile la disconnessione della parte residua del grafo in un componente gigante e tanti subgrafi piccoli).

Quindi: un grafo casuale genera una rete "piccolo mondo" con diametri ridotti e connessioni medie <k> limitate. **Domande che sorgono:**

1. possiamo fare di meglio?
2. i grafi casuali spiegano **tutte** le proprietà delle reti "piccolo mondo"?

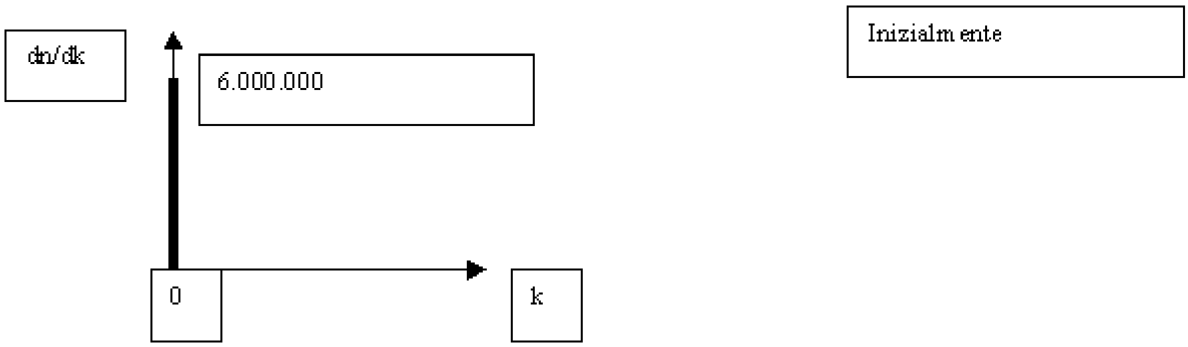
possiamo fare di meglio?

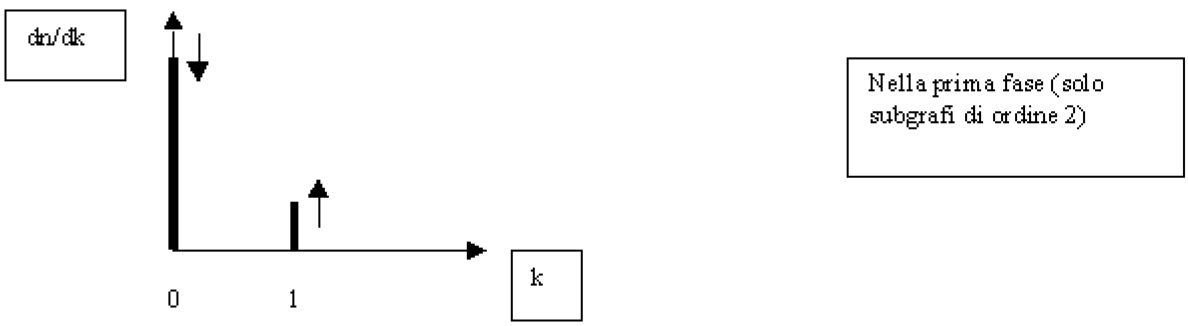
Innanzitutto: definizione di "meglio". Stella (albero a un livello) con k(radice) = 6.000.000.000 e k (foglie) = 1 dà diametro 2 e <k>=2, MA k(max) = 6.000.000.000.

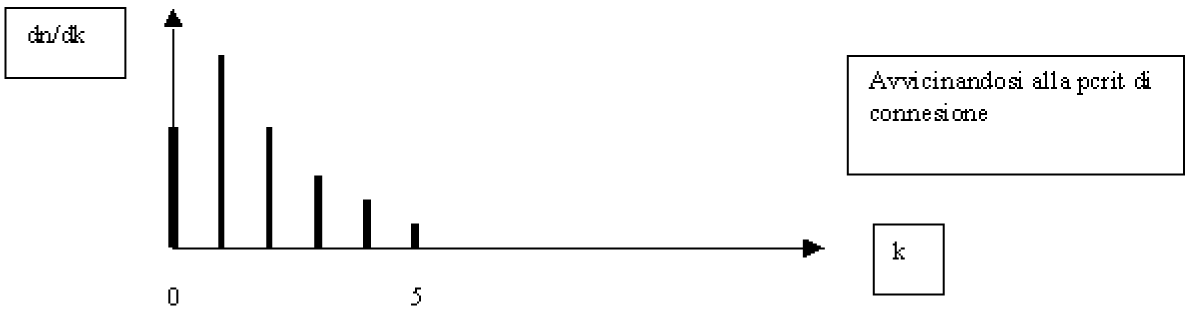
Caratteristiche desiderabili in una rete - concetto di costo di una rete (c = cammino:)

<k> <c>	se k e c sono distribuiti uniformemente o con Poisson
k(max) c(max)	se k e c hanno distribuzioni molto irregolari

Evoluzione della distribuzione di k sulla rete casuale:







(vedi Statistical Mechanics of Complex Networks, pg 12)

Ricerca di reti con k distribuito più uniformemente: si prendono in esame reti ad alto grado di regolarità

Alberi regolari d-ari di profondità p (d=numero di diramazioni ad ogni livello):

N. foglie = d^p (nodi con k = 1)
N. nodi intermedi (p -> ∞) = N. Foglie / (d - 1) (nodi con k = d + 1)
k (max) = d + 1
<k> = 2 (distribuzione di k? Problema interessante per chi voglia affrontarlo)
diametro = 2 p (distribuzione di c? Problema interessante per chi voglia affrontarlo)

Albero				
Nodi	6.000.000.000			
Diramazioni	Profondità circa	Diametro=2*profondità		Costo
			Kmax=Diram+1	
2	33	64	3	192
3	21	40	4	160
4	17	32	5	160
5	14	26	6	156
6	13	24	7	168
7	12	22	8	176
8	11	20	9	180
9	11	20	10	200
10	10	18	11	198
15	9	16	16	256
20	8	14	21	294
25	7	12	26	312
30	7	12	31	372
35	7	12	36	432
40	7	12	41	492
45	6	10	46	460
50	6	10	51	510

Ipercubi generalizzati

Ipercubi generalizzati completi				
Nodi	6.000.000.000			
Base	Dimensioni	Diametro=dim		Costo
			<k>=k(costante)	
2	33	33	33	1089
3	21	21	42	882
4	17	17	51	867
5	14	14	56	784
6	13	13	65	845
7	12	12	72	864
8	11	11	77	847
9	11	11	88	968
10	10	10	90	900
15	9	9	126	1134
20	8	8	152	1216
25	7	7	168	1176
30	7	7	203	1421
35	7	7	238	1666
40	7	7	273	1911
45	6	6	264	1584
50	6	6	294	1764

Ipercubi generalizzati NON completi (reticoli a barre)				
Nodi	6.000.000.000			
Base	Dimensioni	Diametro=(b-1) *d		Costo
			Kmax=*2Dim	
2	33	33	66	2178
3	21	42	42	1764
4	17	51	34	1734
5	14	56	28	1568
6	13	65	26	1690
7	12	72	24	1728
8	11	77	22	1694
9	11	88	22	1936
10	10	90	20	1800
15	9	126	18	2268
20	8	152	16	2432
25	7	168	14	2352
30	7	203	14	2842
35	7	238	14	3332
40	7	273	14	3822
45	6	264	12	3168
50	6	294	12	3528

La rete casuale, con il suo costo minimo = 23 * 7 = 150 ca, resta imbattuta.

Gli alberi le si avvicinano, ma con diametri sempre molto maggiori.

I grafi ad alta regolarità veramente soddisfano altri requisiti delle reti, come la simmetria => facile algoritmo di routing.

29.4.2004

I grafi casuali spiegano tutte le proprietà delle reti "piccolo mondo"?

No. Dati empirici molto diversi per

- coefficente di aggregazione (per reti piccolo mondo varia fra 0.3 e 0.75)
- distribuzione di k (vedremo meglio in seguito)

Definizione di C, coefficente di aggregazione - vedi Collective dynamics of'small-world' di networks di Duncan J. Watts* & Steven H. Strogatz, legenda alla fig. 2

coefficente di aggregazione di una rete casuale = p (densità dei collegamenti), estremamente bassa rispetto a 0.5

Coefficente di aggregazione della rete sociale mondiale? Decisamente superiore, pari a quelle delle reti piccolo mondo. Vedi in merito (sola darci un' occhiata) THE STRENGTH OF WEAK TIES: A NETWORK THEORY REVISITED di Mark Granovetter

Possibile modello che spiega Diametro piccolo e C alto:
Collective dynamics of'small-world' di networks di Duncan J. Watts* & Steven H. Strogatz, (leggerlo tutto, sapere tutto eccetto l' ultima pagina)

Spiegato anche meglio in
Reka Albert1 and Albert-Laszlo Barabasi Statistical Mechanics of Complex Networks
pgg 23 - 26, saltare "average path lenght" e "spectral properties"

Dopo il 5.5.2004 (lezione saltata)

Punti notevoli:

W & S non disponevano di alcuna formula analitica per calcolare C(p) e diametro(p);

solo in seguito altri ricercatori hanno determinato che dev' essere circa

C(p) = ~ C(0) * (1-p) ^ 3
(Statistical Mechanics of Complex Networks, pg 26)
Diam(p) = ~ > Diam (0) / sqrt (2 * p * <k> * Nnodi) (per p lontane da 1)
(**Esempio dei grafi di shortcut completi**): posto nc = n. nodi che partecipano ai shortcut, Nc numero di corrispondenti connessioni "casuali", è
diam (Nc) = ~ 1 + diam (0) / nc
con Nc = nc(nc-1)/2 =~ nc*nc/2 nc = sqrt(2*Nc)

==> intuizione, verifica solo numerica con selezione casi via Montecarlo;

Il plot di C(p) e Diam (p) su scala lineare non dà indicazioni immediate: le due curve precipitano entrambe a valori molto bassi per poi mantenersi basse fino a p=1, a una prima vista potevano convalidare l' ipotesi che D e Diam procedessero di conserva, alti C corrispondenti ad alti Diam e viceversa; le differenza è data dagli andamenti per p basse delle due curve

Per apprezzare la differenza bisognava effettuare uno **zoom sui p bassi**, cosa che risulta automatica utilizzando per p una scala logaritmica x=Log p p = 10^x (p=0 va fuori scala, p=1 corrisponde a x=0) il plot si fa per valori negativi di x 0, -1, -2 ... corrispondenti a p= 1, 0.1, 0.001 eccetera. In questo caso, y rimane lineare in C e Diam

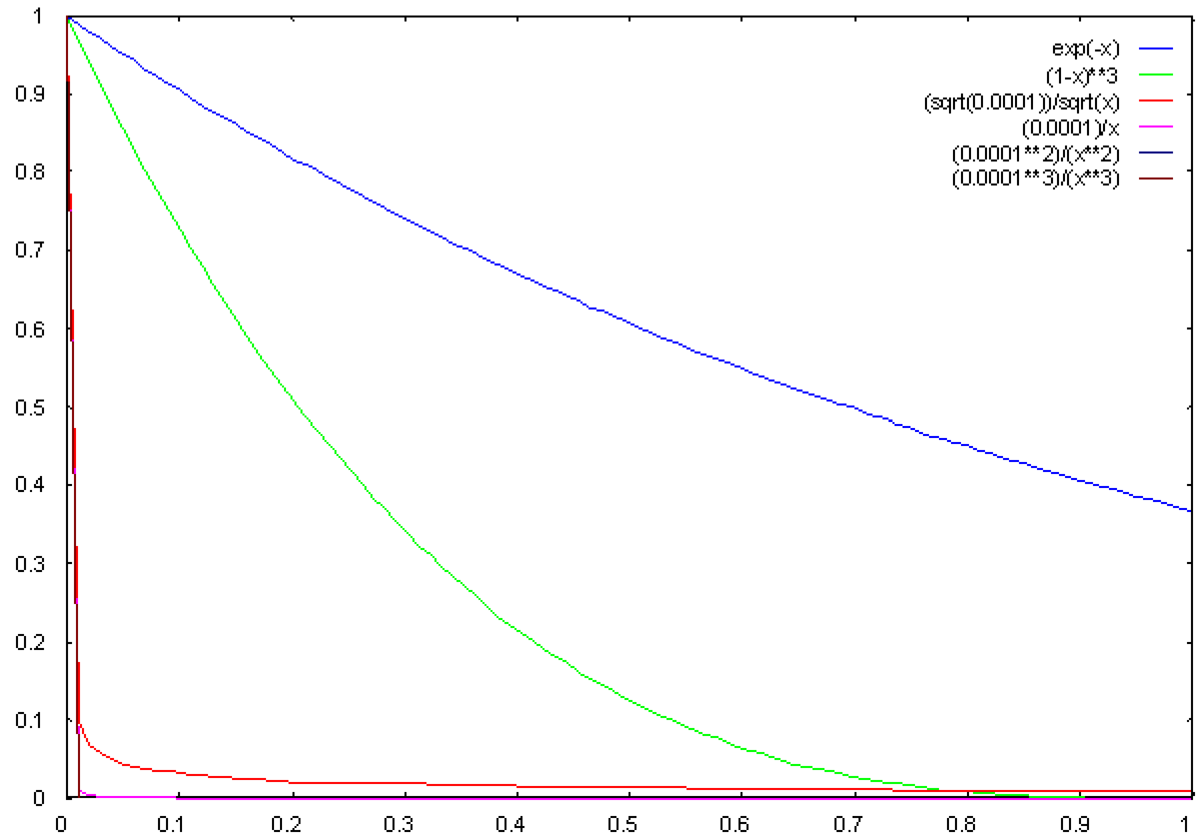
Passando da scala lineare a scala logaritmica molte cose appaiono più chiare.

Cosa accade ponendo anche y=Log C/C(0) e y=Log Diam/Diam(0) ?

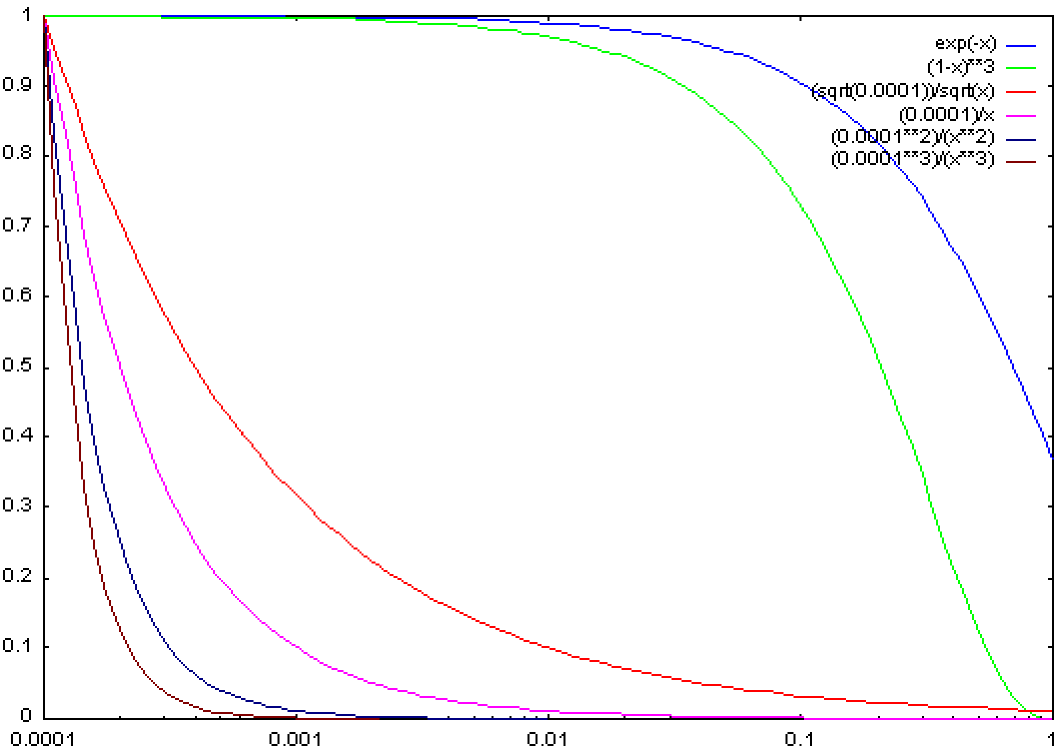
Si Zooma anche sulla parte bassa di C e Diam; ln(Diam(ln p)) diventa una retta discendente con pendenza 1/2; l' altra curva rimane con curvatura verso il basso perché non è in p ma in (1-p).

Vediamo alcune tipiche "curve di decadimento" in scala lineare, x logaritmica, x e y logaritmiche:

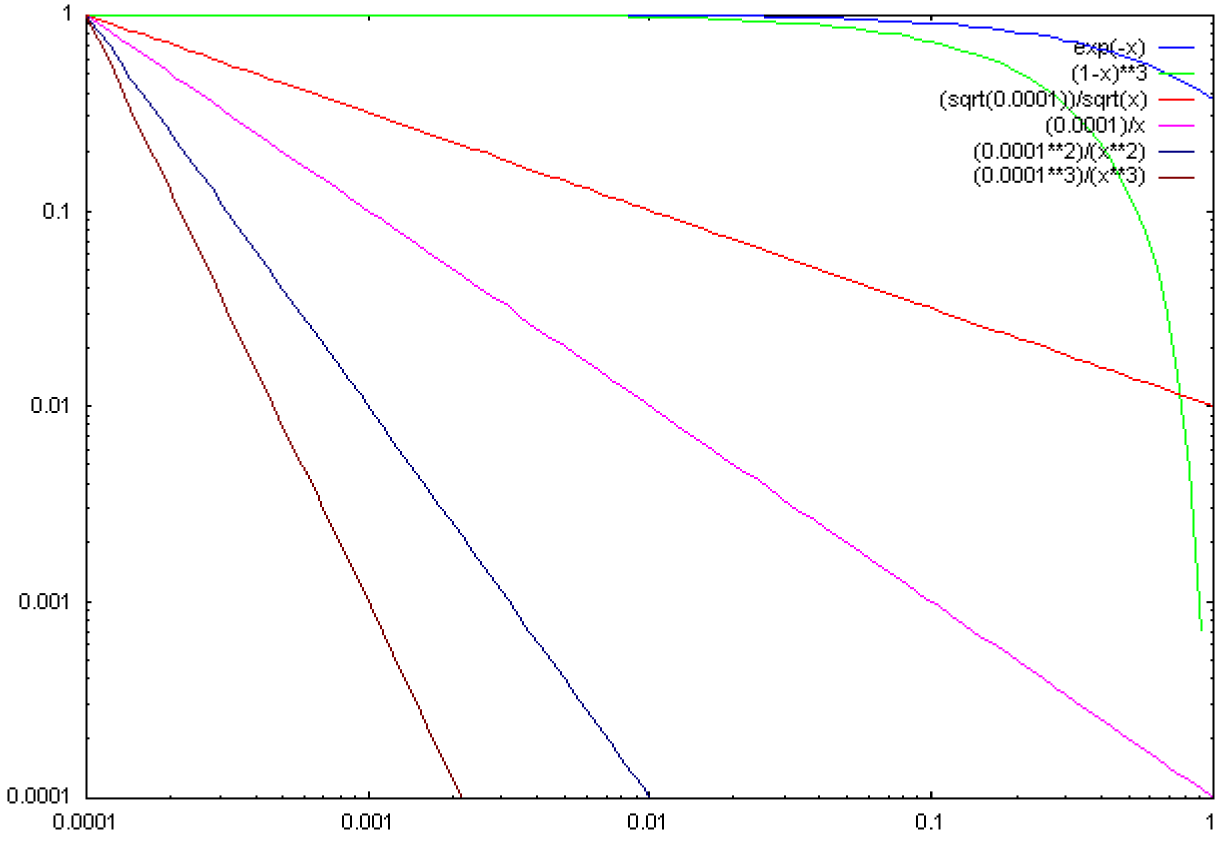
```
gnuplot> set xrange [0.0001:1]
gnuplot> set yrange [0.0001:1]
gnuplot> plot exp(-x), (1-x)**3, (sqrt(0.0001))/sqrt(x), (0.0001)/x, (0.0001**2)
/(x**2), (0.0001**3)/(x**3)
```



```
gnuplot> set logscale x
gnuplot> plot exp(-x), (1-x)**3, (sqrt(0.0001))/sqrt(x), (0.0001)/x, (0.0001**2)
/(x**2), (0.0001**3)/(x**3)
```

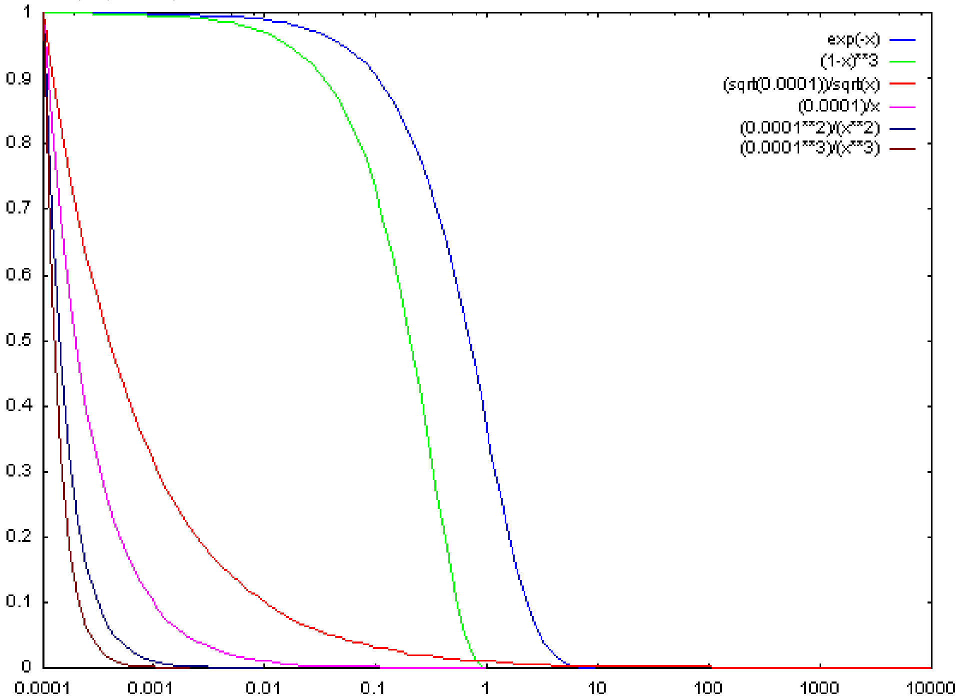



```
gnuplot> set logscale
gnuplot> plot exp(-x), (1-x)**3, (sqrt(0.0001))/sqrt(x), (0.0001)/x, (0.0001**2)
/(x**2), (0.0001**3)/(x**3)
```

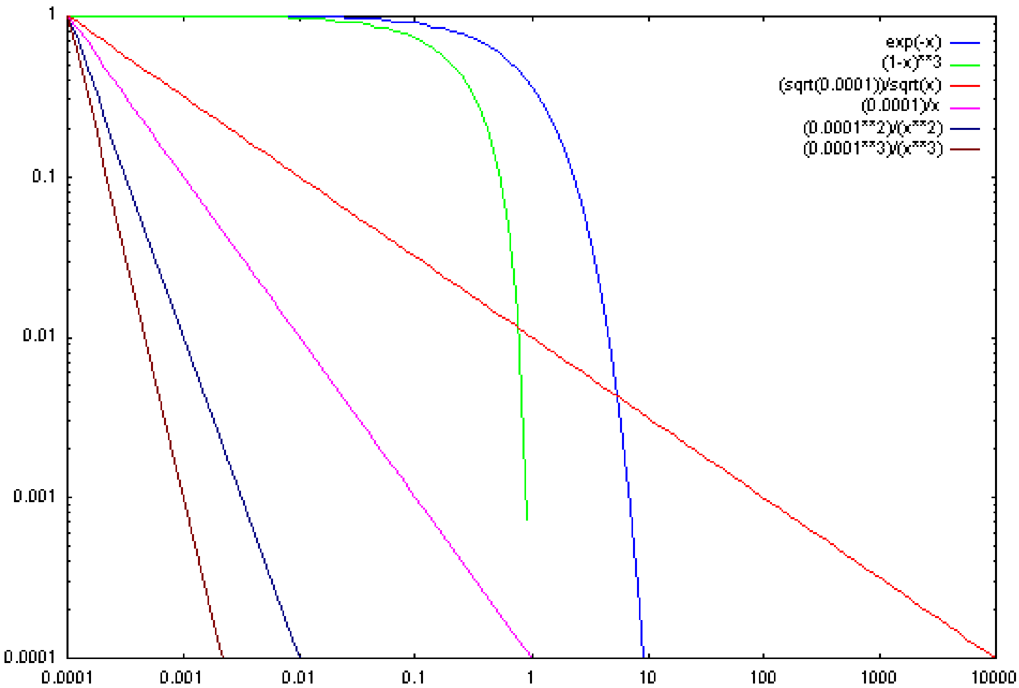


Decadimenti per x >> 1:

gnuplot> set xrange [0.0001:10000]
gnuplot> set logscale x
gnuplot> plot exp(-x), (1-x)**3, (sqrt(0.0001))/sqrt(x), (0.0001)/x, (0.0001**2)/(x**2), (0.0001**3)/(x**3)



gnuplot> set logscale
gnuplot> plot exp(-x), (1-x)**3, (sqrt(0.0001))/sqrt(x), (0.0001)/x, (0.0001**2)/(x**2), (0.0001**3)/(x**3)



Un altro punto di vista sui grafi di Watts e Strogatz:

si possono considerare cluster locali di (<k> - c) nodi totalmente connessi; il problema diventa connettere con una rete casuale i cluster e non i nodi, sfruttando le “c” connessioni di ogni nodo non interne al suo gruppo.

Cambiano le dimensioni: N diventa N/(<k>-c); ma ogni connessione a un gruppo può essere effettuata ad uno qualunque dei suoi nodi, e quindi aumenta il grado del gruppo solamente di 1/(<k>-c)

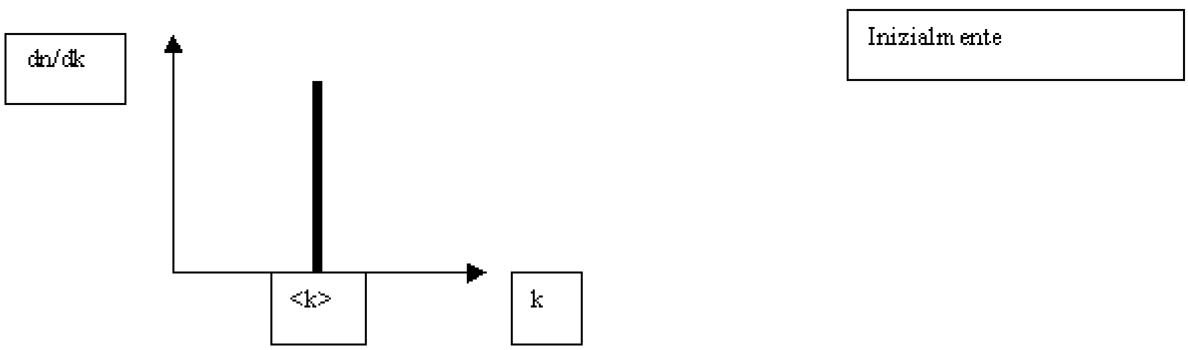
Per N = 6.000.000.000
<k> = 23 (quella della probabilità critica di connessione)
c = 3; quindi nodi in un gruppo = 20
N gruppi 300.000.000
<k gruppi> di probabilità critica di connessione -> 19,5...
Diam = Diam (kcrit gruppi) + 1 + 1 = 6,6 + 2 = 8,6

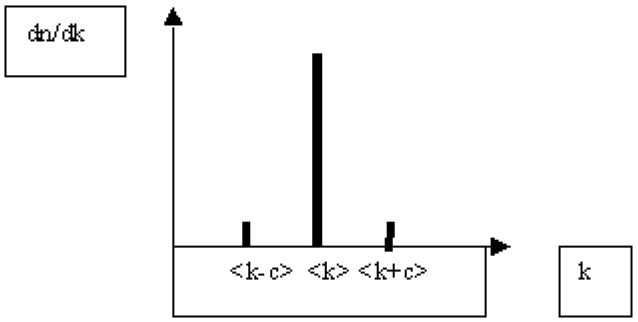
ma qui ci avanzano c * (k-c) = 60 archi per ogni gruppo...
Diam = Diam (3 * kcrit gruppi) + 1 + 1 = 4,8 + 2 = 6,8

Toh, i 6 - 7 gradi di separazione...

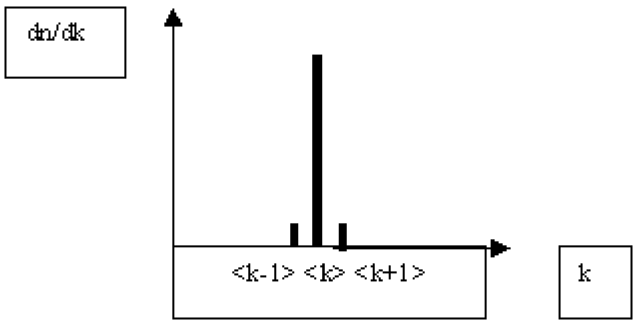
Distribuzioni di K

Evoluzione della distribuzione di k sulla rete ordinata con introduzione di connessioni casuali:

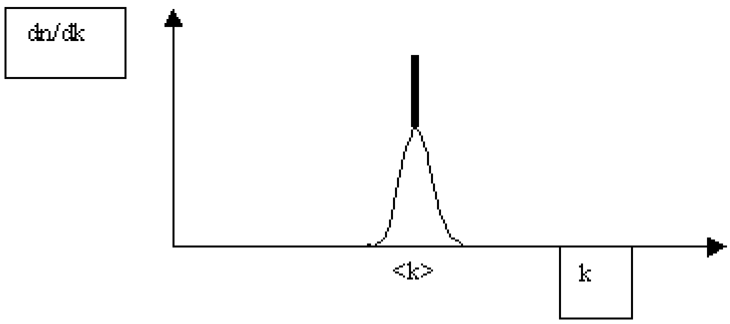




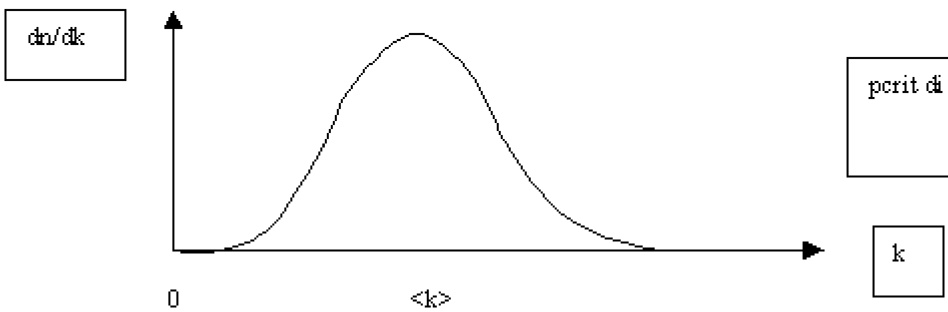
shortcuts di grafi completi



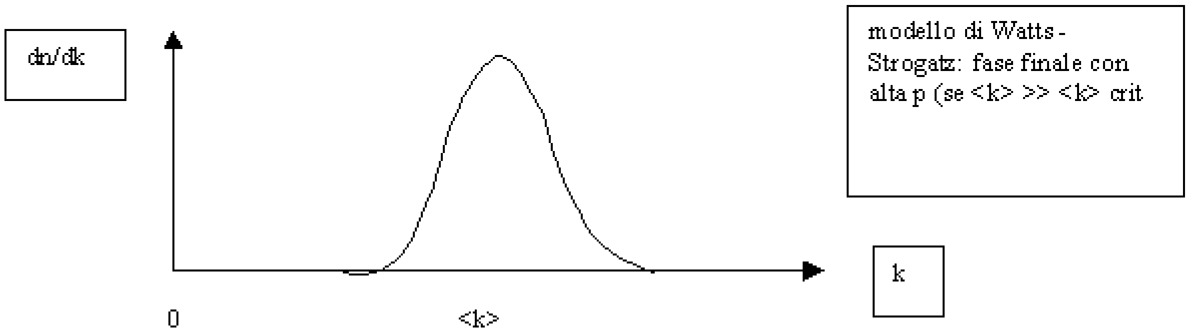
modello di Watts-Strogatz: fase iniziale



modello di Watts-Strogatz: fase finale con bassa p



pcrit di connesione



Distribuzione di k nelle reti reali...

....