

Функциональные сети: структура, динамика, ЭВОЛЮЦИЯ

к.ф.м.-н. И.И. ТИТОВ

«Последовательность - лишь первый уровень понимания генома» Venter et al.

Следующий этап постгеномной эры - исследование взаимодействий

- между генами

- между их продуктами

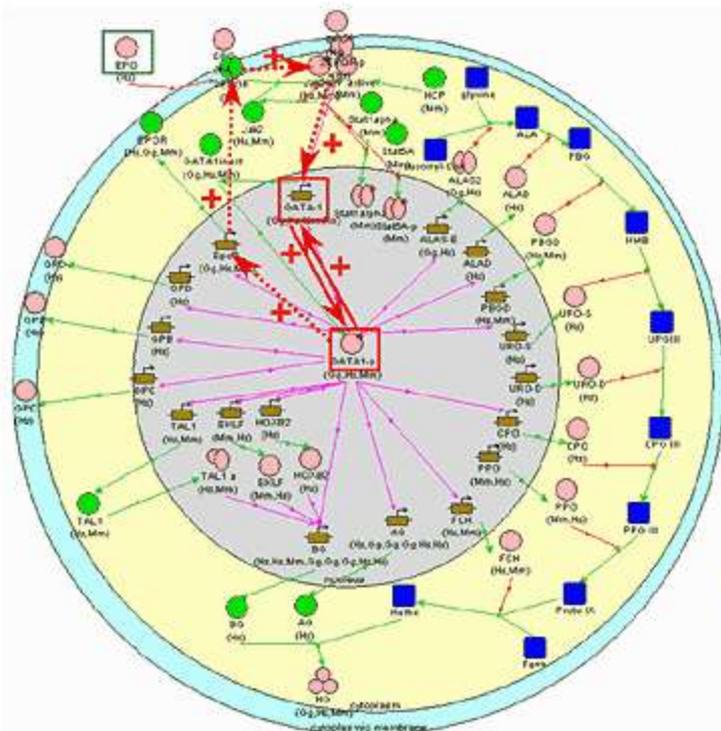
Индуктивные и репрессивные взаимодействия

Какие бывают сетевые взаимодействия?

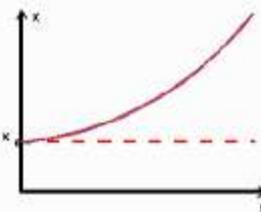
Положительная обратная связь

БАЗА ДАННЫХ GENENET:

ФРАГМЕНТ ГЕННОЙ СЕТИ ДИФФЕРЕНЦИРОВКИ И СОЗРЕВАНИЯ ЭРИТРОЦИТОВ



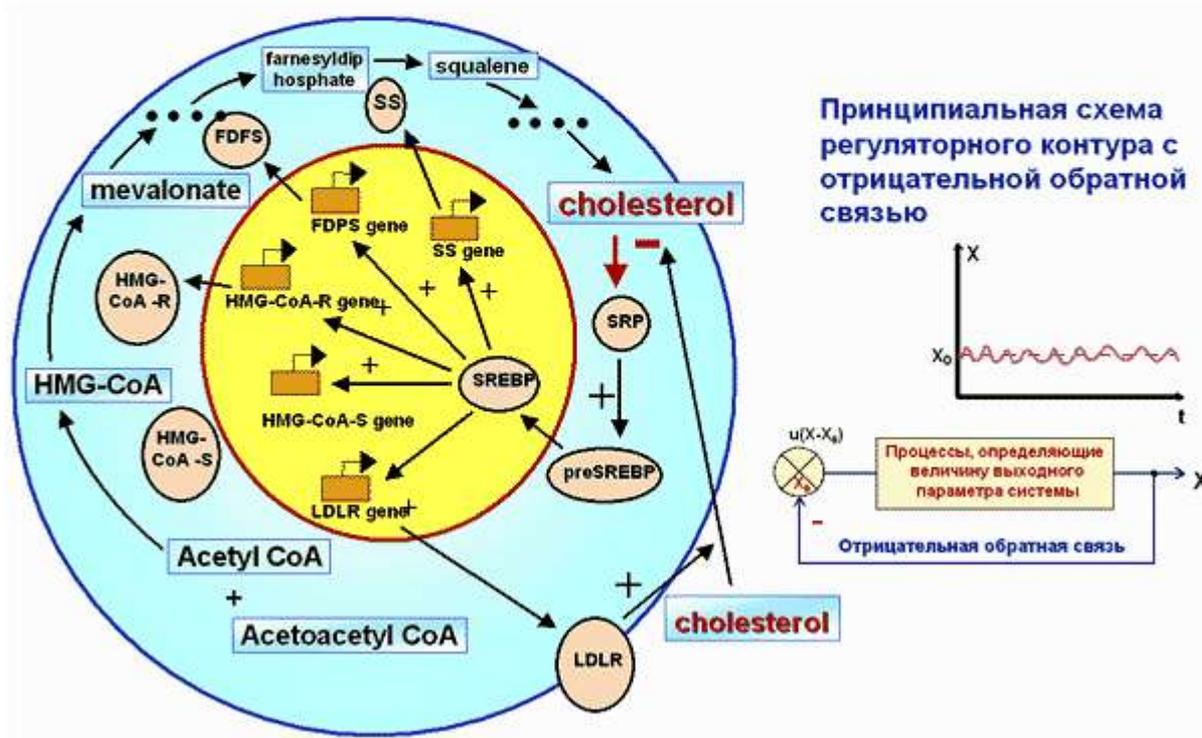
$$\frac{\partial x}{\partial t} \propto x$$



отрицательная обратная связь

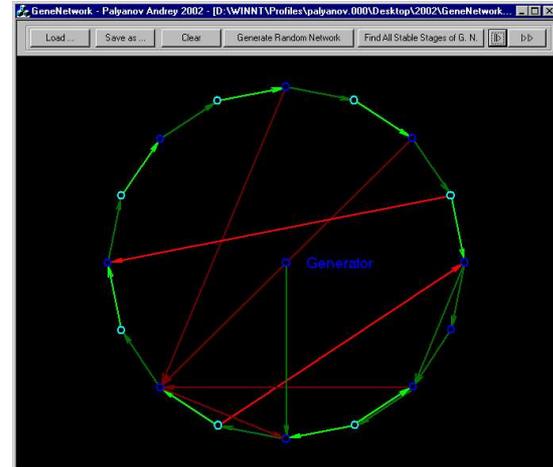
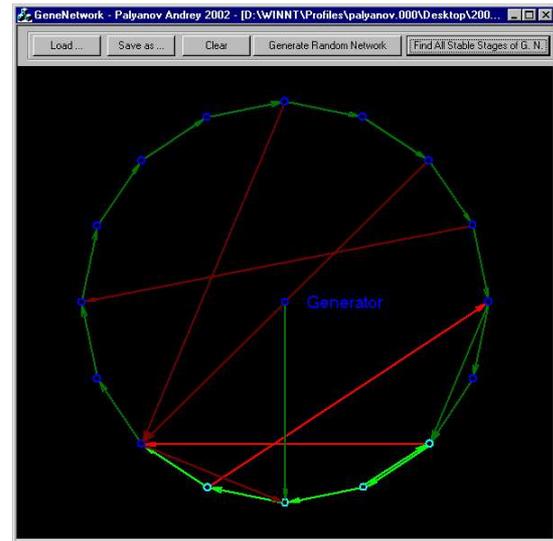
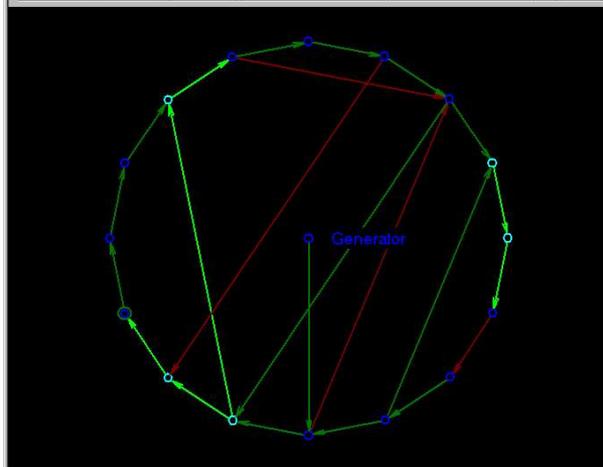
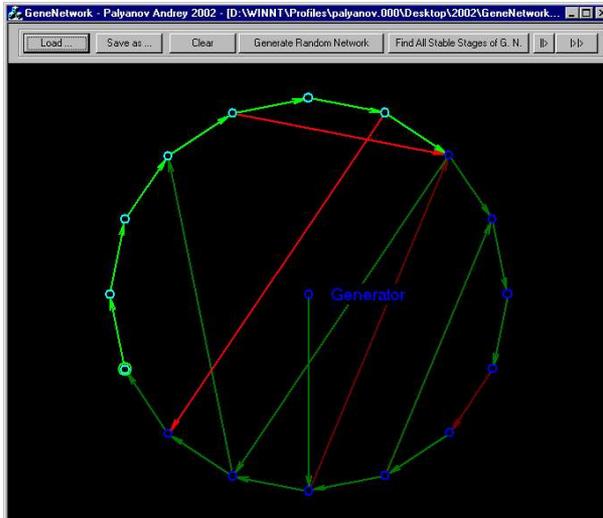
БАЗА ДАННЫХ GENENET:

ЦЕНТРАЛЬНЫЙ ФРАГМЕНТ ГЕННОЙ СЕТИ БИОСИНТЕЗА ХОЛЕСТЕРИНА В КЛЕТКЕ (РЕГУЛЯЦИЯ ПО МЕХАНИЗМУ ОТРИЦАТЕЛЬНОЙ ОБРАТНОЙ СВЯЗИ)

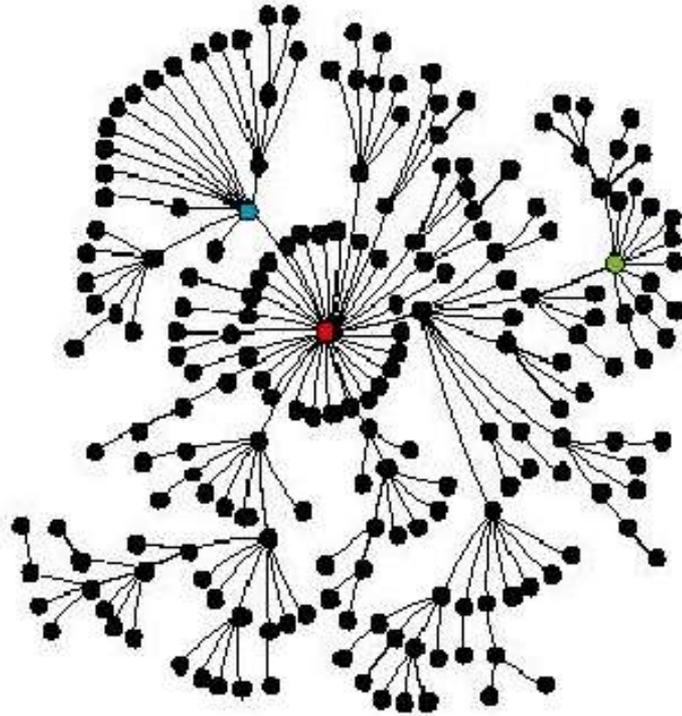


От одиночного к сети взаимодействий: виды активации

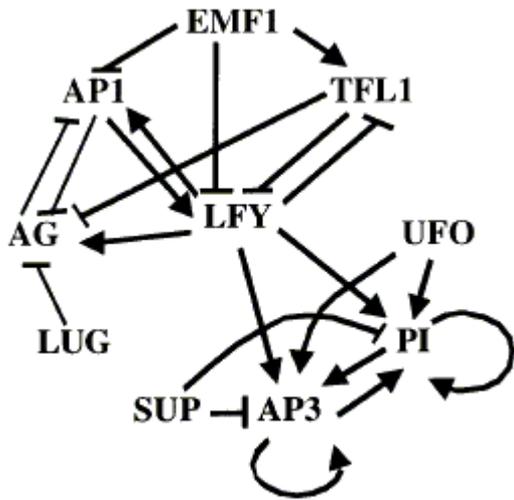
циклическая



Кассетная активация



Наблюдаемое vs существующее; переключения режимов

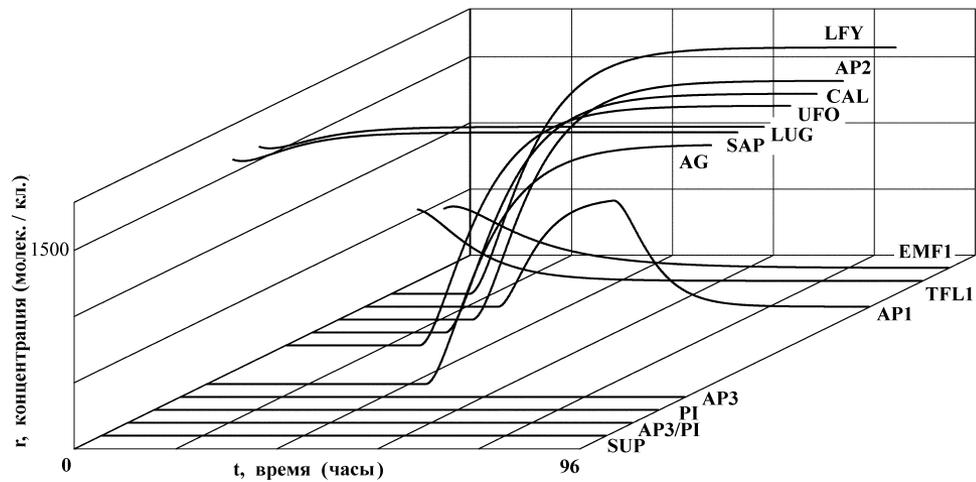
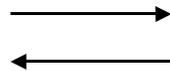
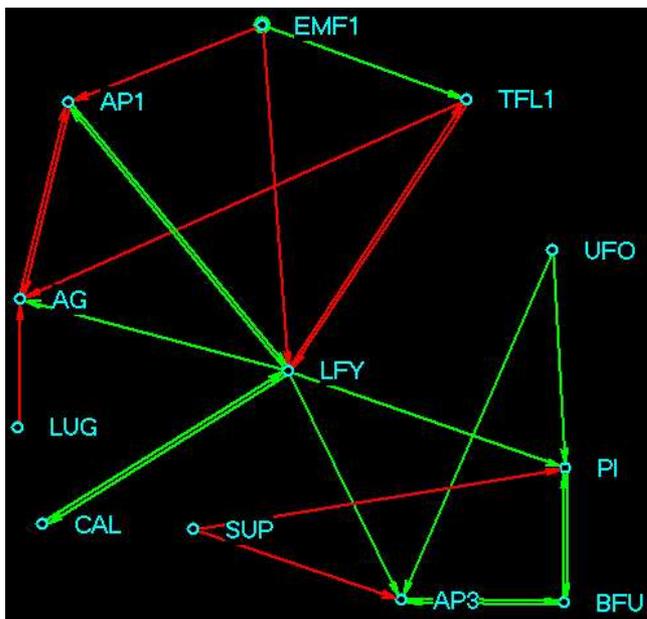


1. Прямая задача для генных сетей

Дана структура и параметры - найти поведение

2. Обратная задача

дано поведение - найти структуру и параметры



Основные структурные характеристики узлов и сетей

Степень вершины - число соединений с соседями

Геодезическое расстояние - кратчайший путь между вершинами, среднее по сети называется диаметром

Коэффициент кластеризации – вероятность быть соседями для вершин, имеющих общего соседа

Размер гигантской компоненты – число вершин в самом большом кластере

Простейшая модель сети – графы со случайными связями (Эрдеш и Реньи)

Два параметра сети:

N – число узлов

P – вероятность связи

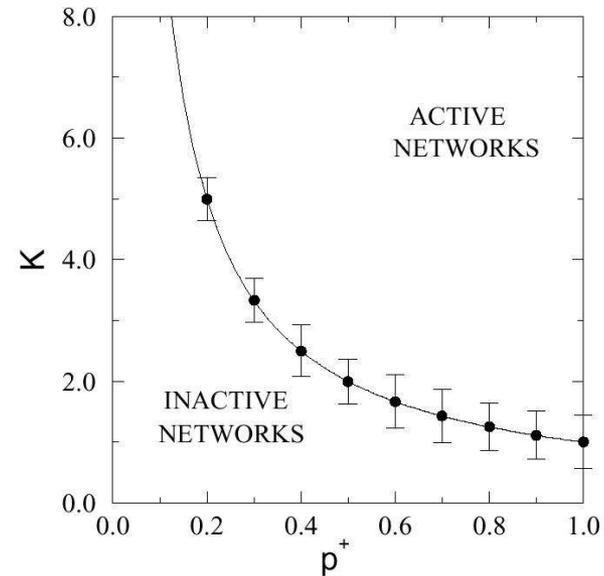
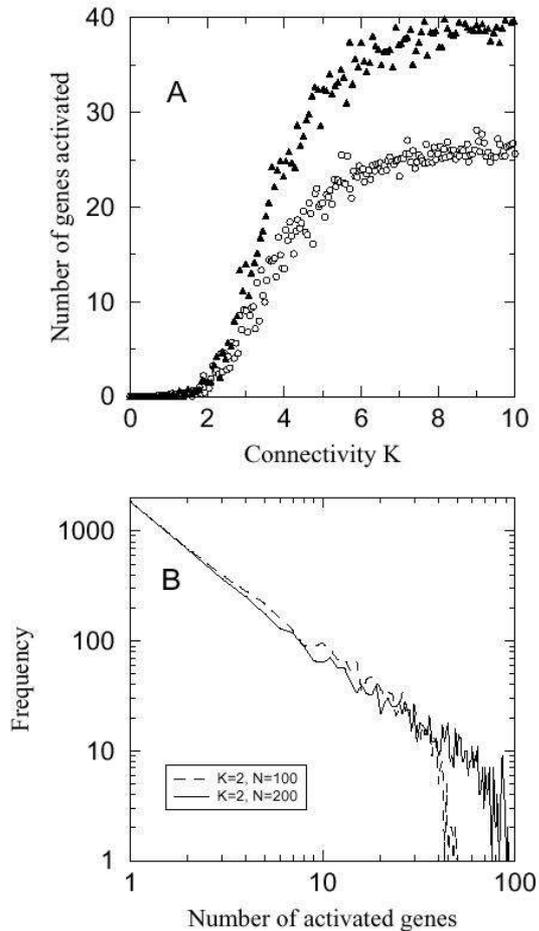
Число связей $pN(N-1)/2$

Биномиальное распределение по числу соседей, среднее число связей на вершину $\langle k \rangle = p(N-1)$

Порог перколяции $\langle k_c \rangle = 1$

Диаметр?

Простейшая модель сети – графы со случайными связями (Эрдеш и Реньи)

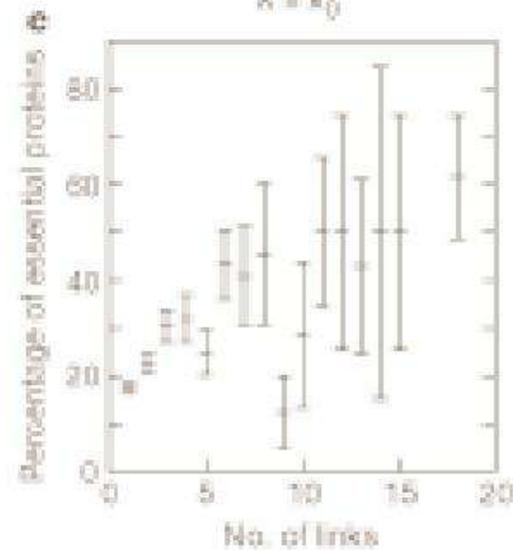
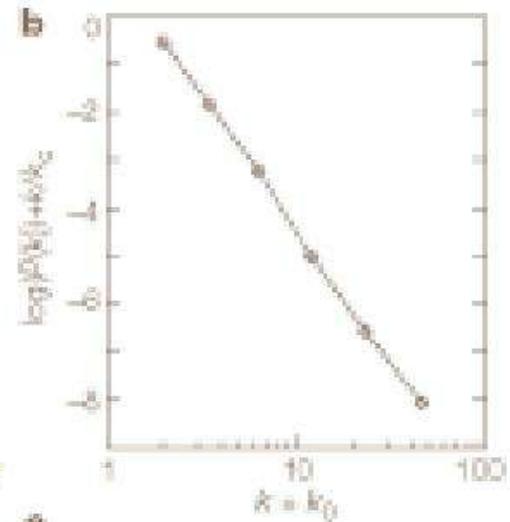


Фазовая диаграмма
случайной геной сети

p - вероятность активации

Фазовый переход в
случайной геной сети

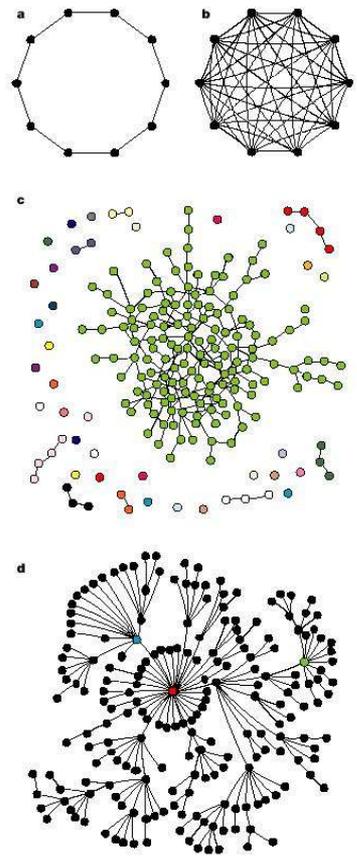
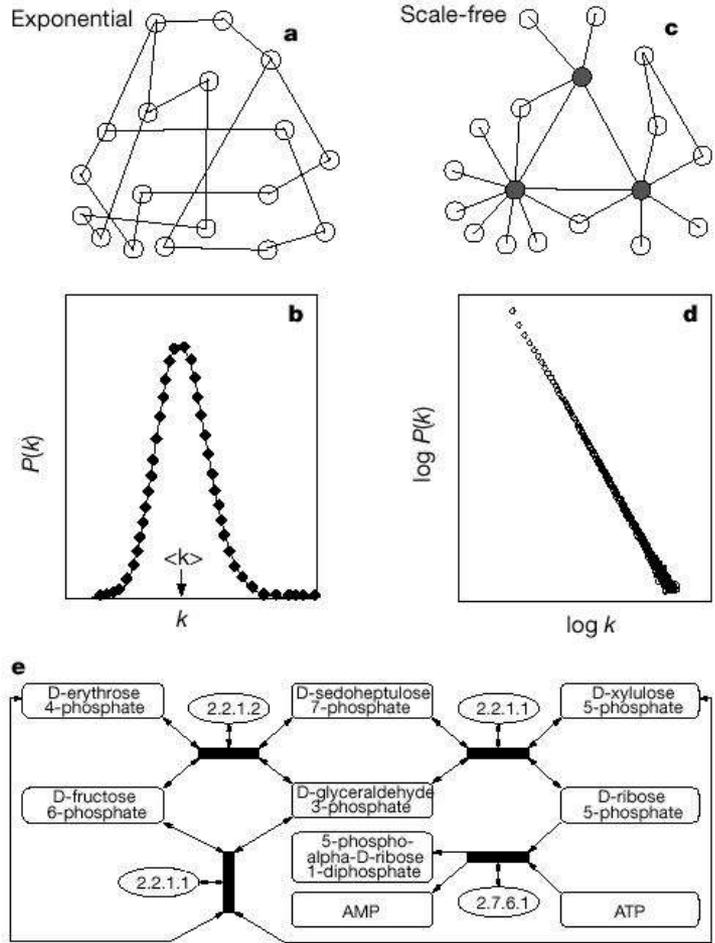
Протеома E.Coli: 1870 белков и 2240 идентифицированных взаимодействий



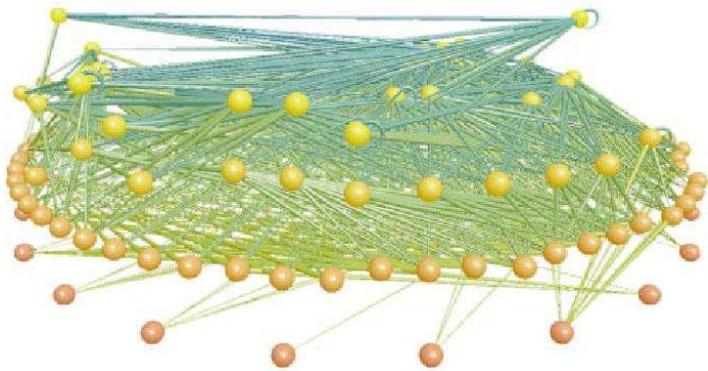
Исследованные безмасштабные сети

- биологические: нейронные, пищевые, экологические, метаболические, генные
- языковые
- социальные: цитирования научных статей, соавторов, содиректоров, актеров, половых контактов (сети дружбы – экспоненциальные, почему?)
- коммуникационные: интернет, www
- радиоэлектронные и электрические
- программного обеспечения
- конформационное пространство макромолекул

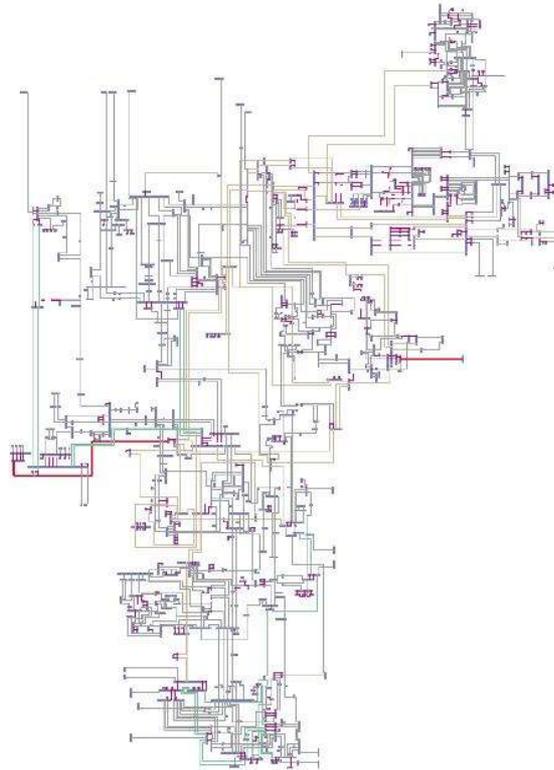
Что такое безразмерная сеть



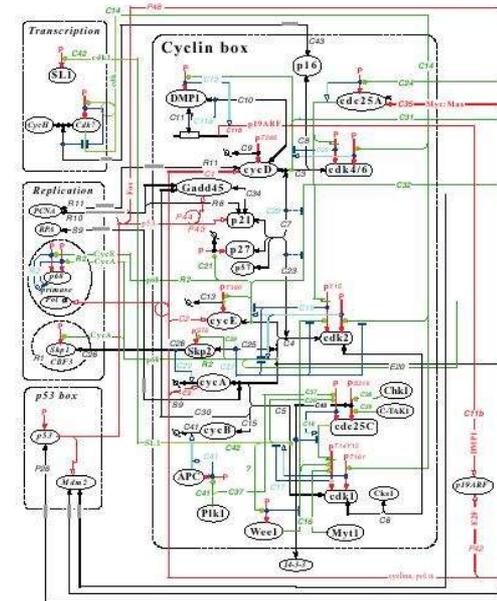
Архитектура регулярных и случайных сетей



Пищевая карта озера
Литтл Рок

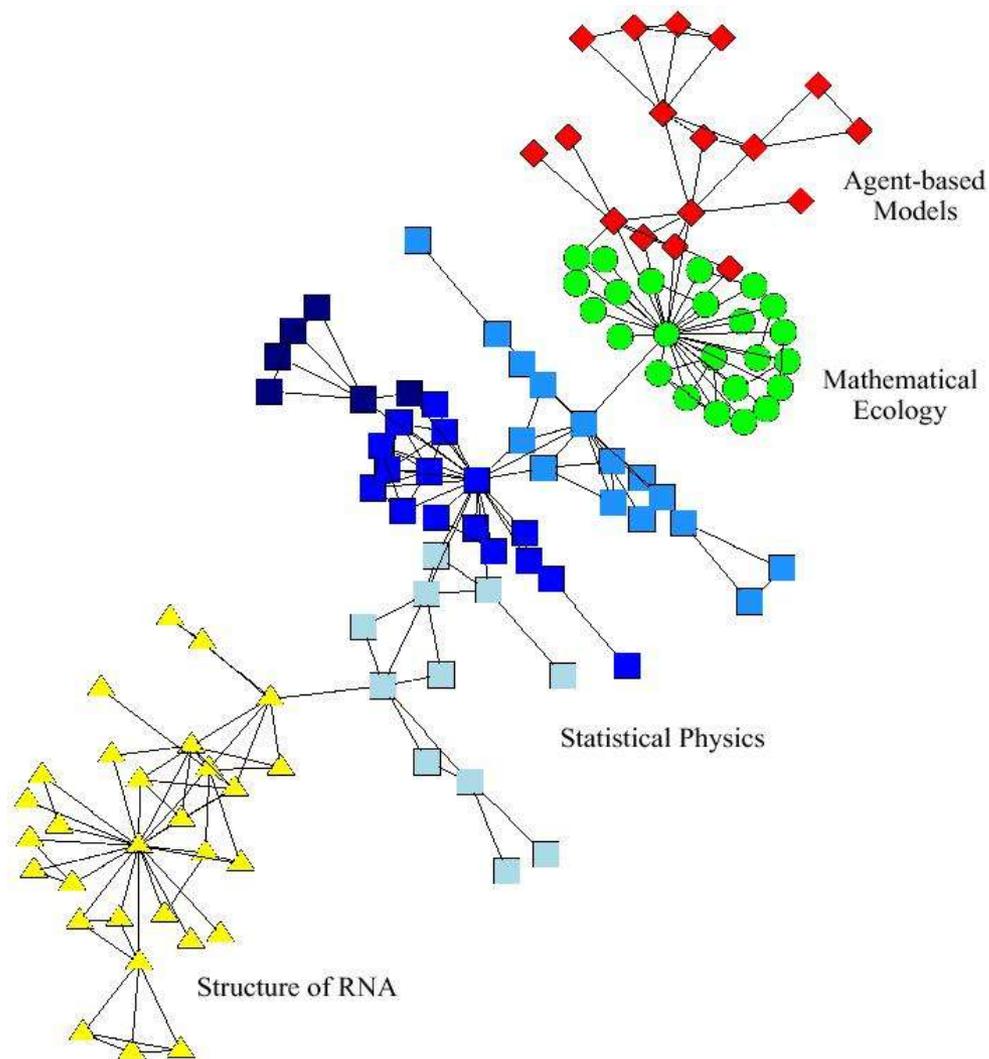


Электросеть штата Нью Йорк



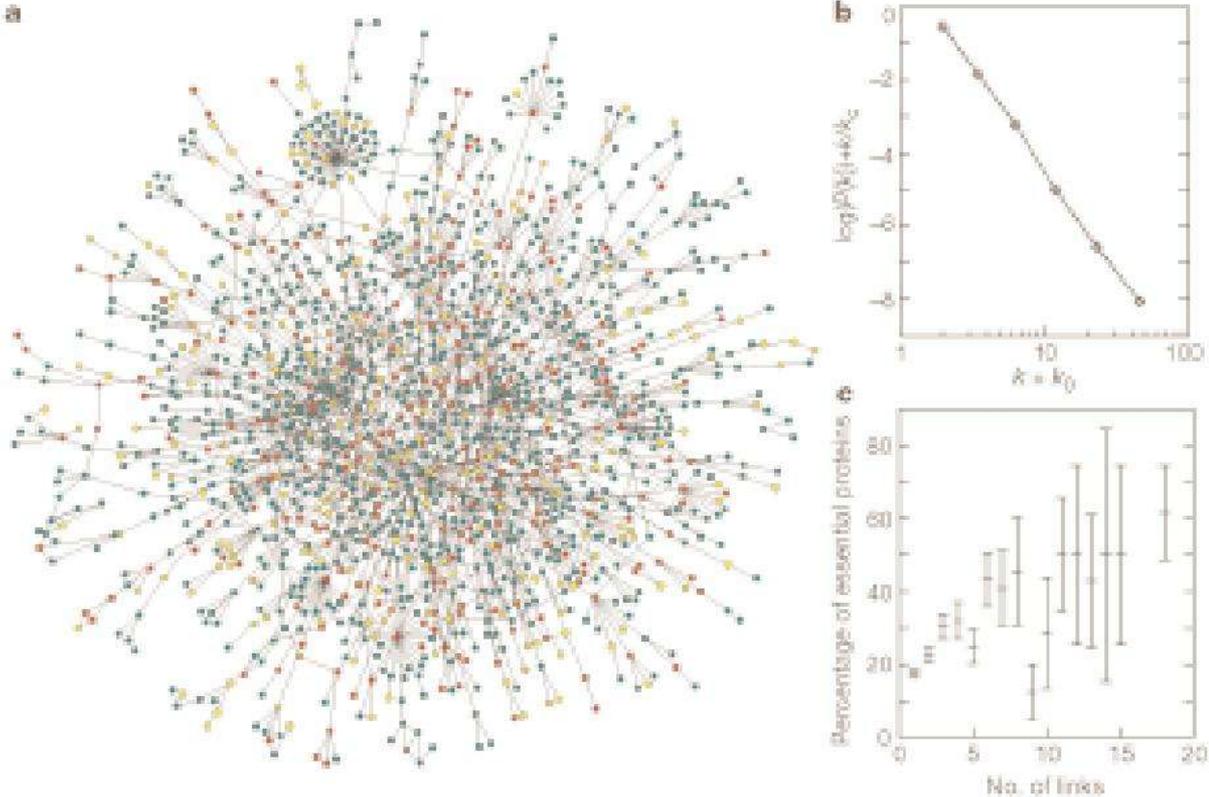
часть карты
молекулярных
взаимодействий
регуляторной сети,
контролирующей
клеточный цикл у
млекопитающих

Главная компонента графа сети взаимодействия научных работников в Институте Санта-Фе



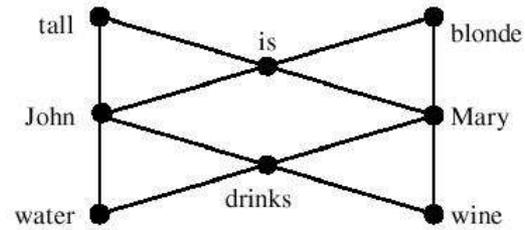
Характеризация протеомы E.Coli

- а) карта бело-белковых взаимодействий; наибольший кластер содержит 78% белков
- б) распределение связности
- в) зависимость числа существенных белков от их связности



Словарные сети

(a)



(b)

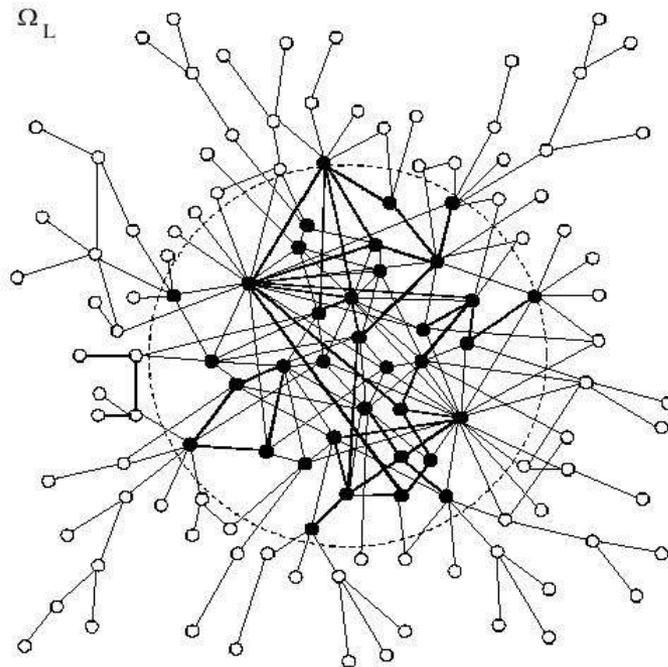
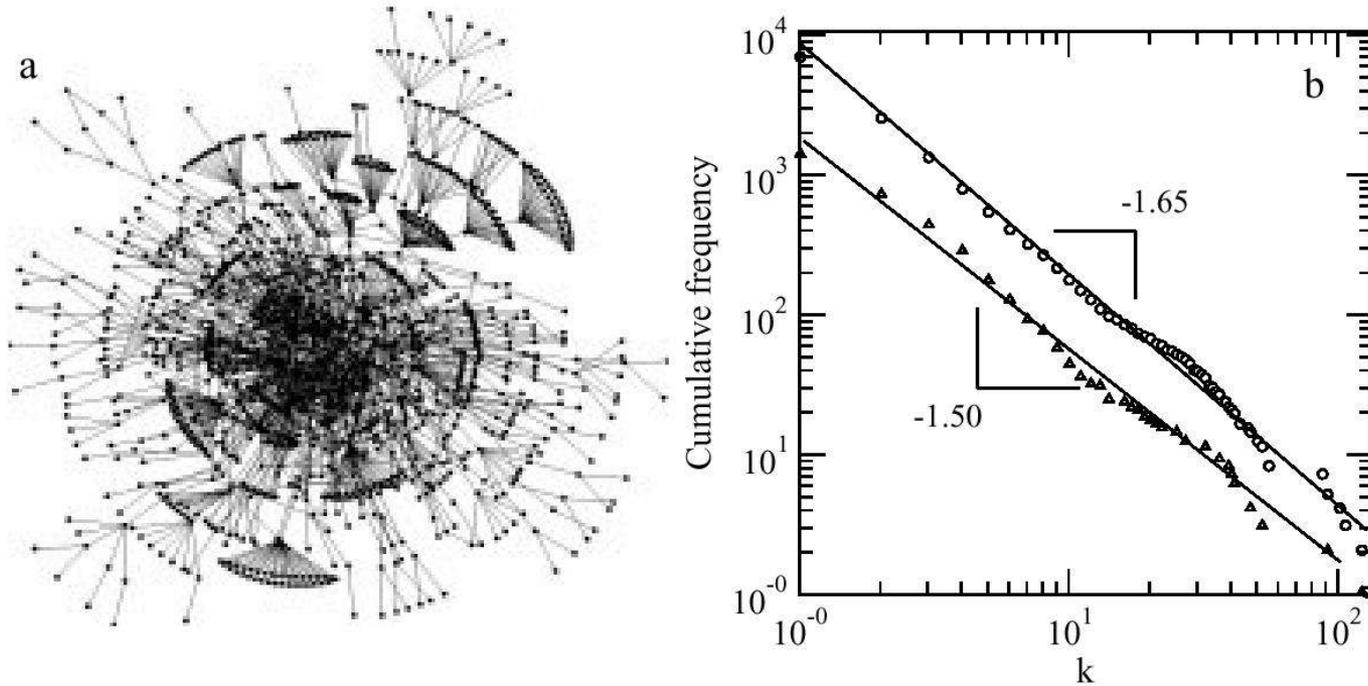


Диаграмма классов Java Development Framework 1.2



Параметры реальных сетей

Network	N	ℓ	C	C_{rand}
movie actors	225 226	3.65	0.79	0.00027
neural network	282	2.65	0.28	0.05
power grid	4941	18.7	0.08	0.0005

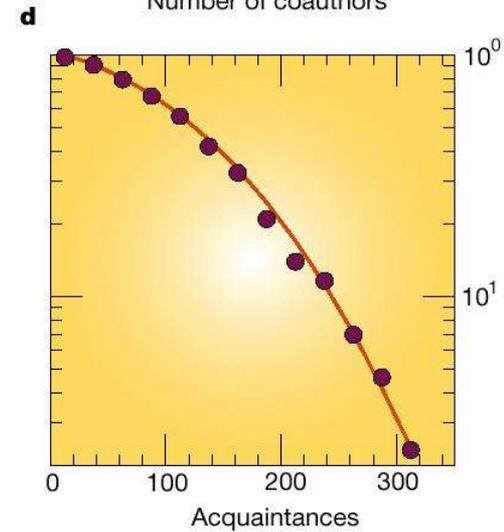
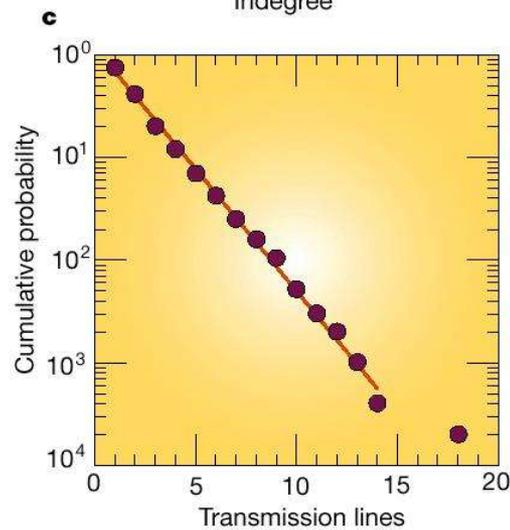
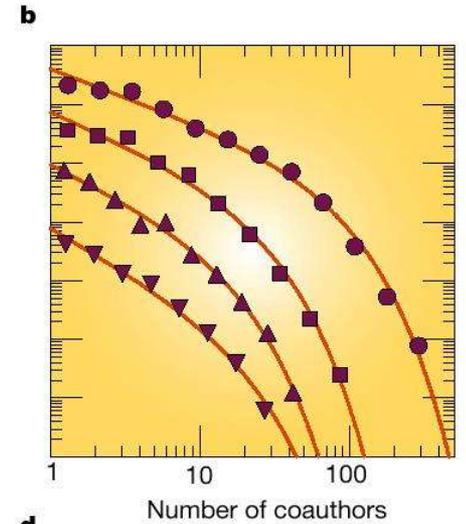
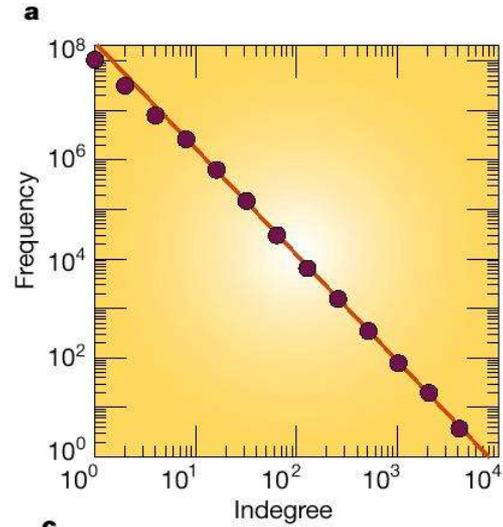
Распределение связности реальных сетей

интернет

соавторство

электросеть США и Канады

социальная сеть мормонов



Циклы в метаболических сетях

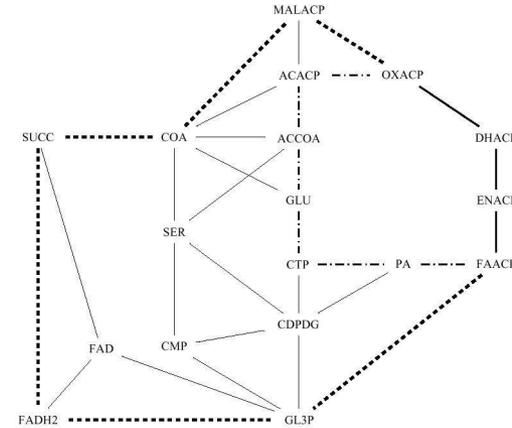
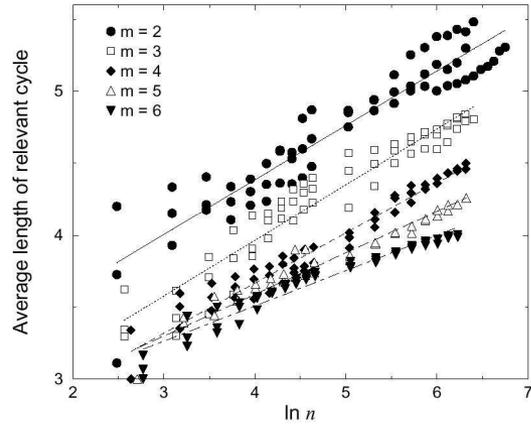


TABLE I: Cycle Structure of Metabolic Networks.

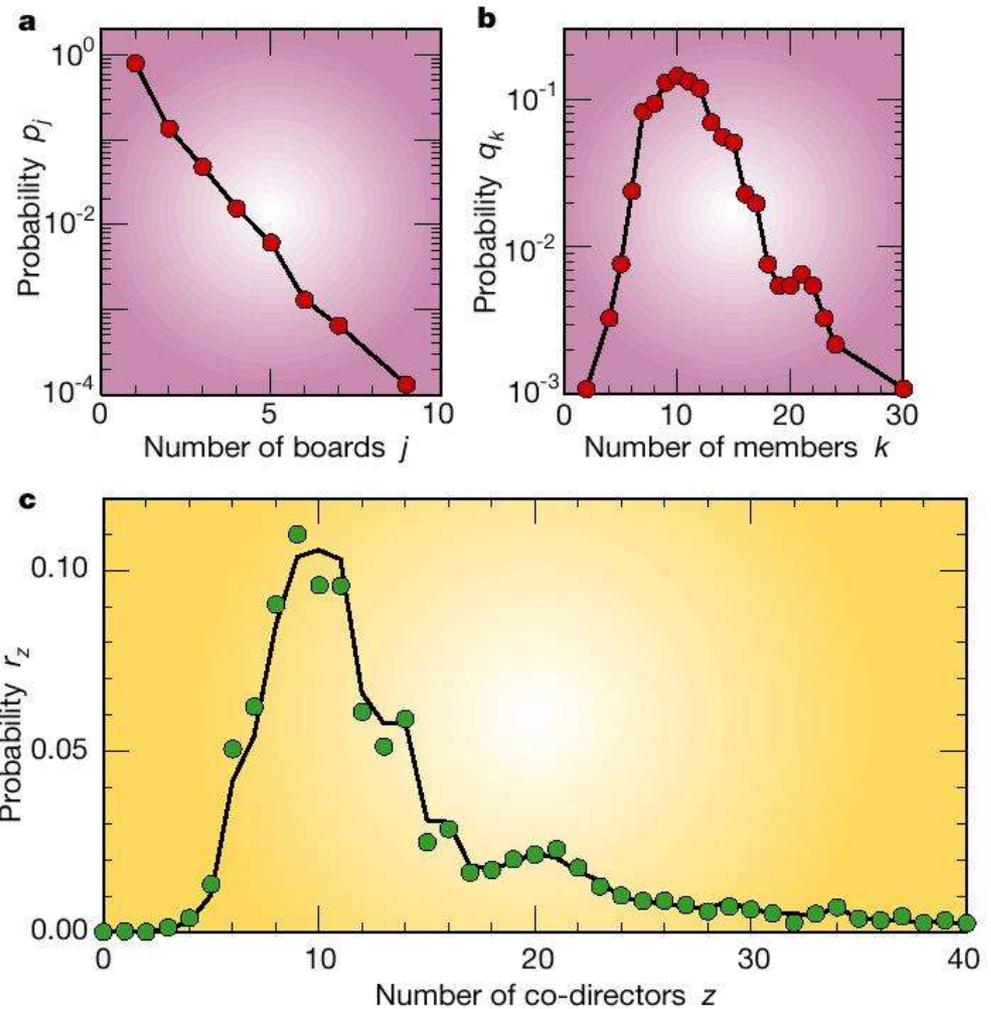
Model	$ C $	3	4	5	6	7	8	9	Σ
Ecoli1	MCB	282	51	19	20	3	5	1	381
	\mathcal{R}	379	114	90	83	5	36	16	723
	\mathcal{S}	379	56	24	42	2	14	16	533
AB	MCB	78	158	124	20	0.4	0.01	0	380
	\mathcal{R}	81	285	527	161	5.5	0.4	0	1060
	\mathcal{S}	81	273	414	144	5.5	0.4	0	918
ER	MCB	18	58	163	131	11	0.4	0	381
	\mathcal{R}	18	61	212	528	82	3.2	0	904
	\mathcal{S}	18	61	205	311	68	3.2	0	666
SW1	MCB	15	46	131	167	21	1.1	0.03	381
	\mathcal{R}	15	48	157	427	151	7.1	0.2	805
	\mathcal{S}	15	48	155	301	108	6.5	0.2	634

Сеть директоров корпораций

распределение числа постов на одного
директора

распределение корпораций по числу
директоров

распределение по числу со-директоров

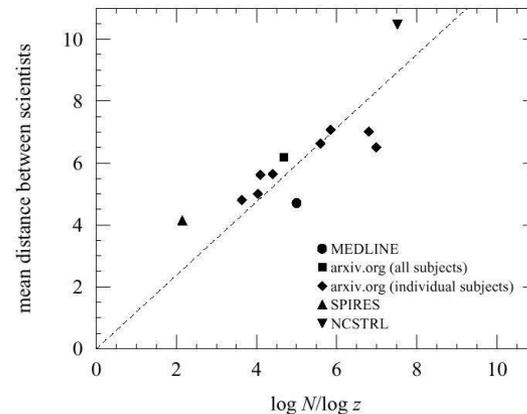
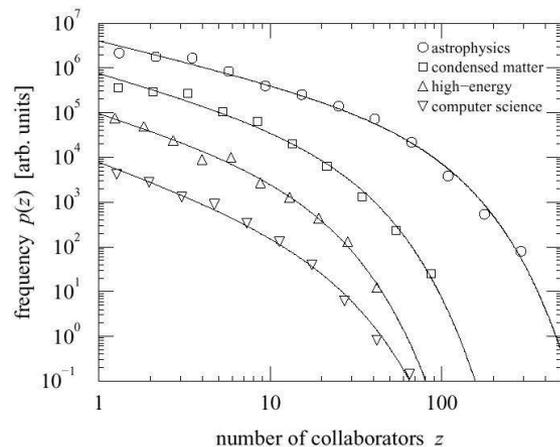


Характеристики кластеризации в трех сетях

Network	Clustering C	
	Theory	Actual
Company directors	0.590	0.588
Movie actors	0.084	0.199
Biomedical authors	0.042	0.088

US corporate directors: 7,673 company directors linked by joint membership on 914 boards of the Fortune 1,000 companies for 1999. Movie actors: 449,913 actors linked by mutual appearances in 151,261 feature films, as specified by the Internet Movie Database (www.imdb.com) as of 1 May 2000. Biomedical collaborations: 1,388,989 scientists linked by coauthorship of at least one of 2,156,769 biomedical journal articles published between 1995 and 1999 inclusive, as listed in the MEDLINE database. The clustering coefficient C is defined as the probability that a connected triple of nodes is actually a triangle; here nodes correspond to people, as in the unipartite representation shown in Fig. 7b. Intuitively, C measures the likelihood that two people who have a mutual collaborator are also collaborators of each other. The results show that the random model accurately predicts C for the corporate director network, given the network's bipartite structure and its degree distributions; no additional social forces need to be invoked. For the networks of actors and scientists, the model accounts for about half of the observed clustering. The remaining portion depends on social mechanisms at work in these communities (see text). (Adapted from ref. 91.)

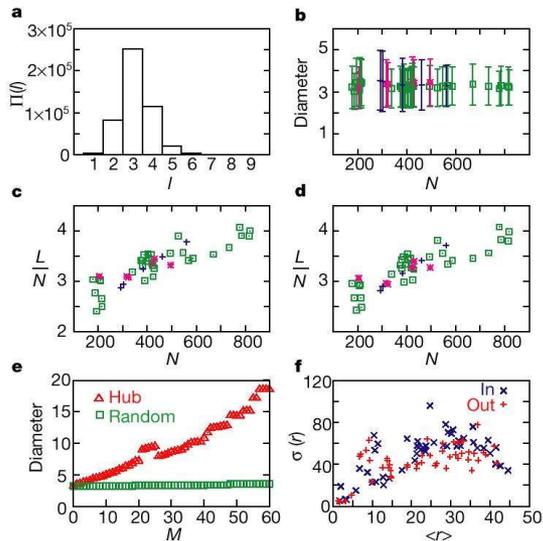
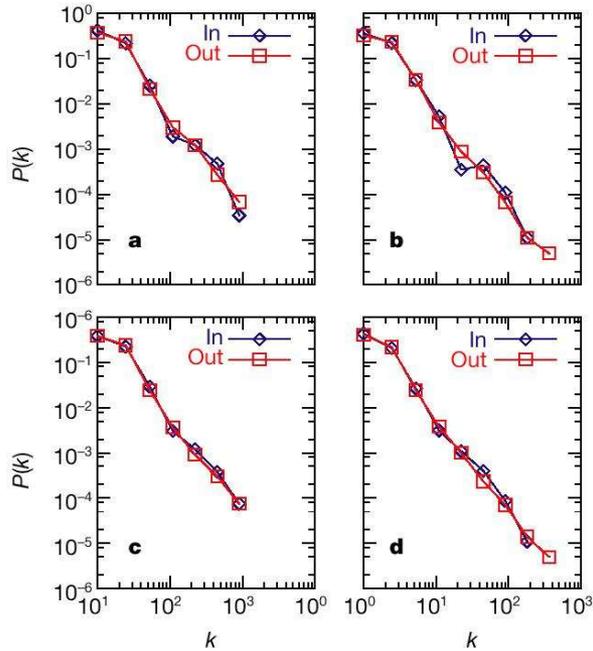
Структура научных сообществ



	MEDLINE	Los Alamos e-Print Archive				SPIRES	NCSTRL
		complete	astro-ph	cond-mat	hep-th		
total papers	2156769	98502	22029	22016	19085	66652	13169
total authors	1388989	52909	16706	16726	8361	56627	11994
first initial only	1006412	45685	14303	15451	7676	47445	10998
mean papers per author	5.5(4)	5.1(2)	4.8(2)	3.65(7)	4.8(1)	11.6(5)	2.55(5)
mean authors per paper	2.966(2)	2.530(7)	3.35(2)	2.66(1)	1.99(1)	8.96(18)	2.22(1)
collaborators per author	14.8(1.1)	9.7(2)	15.1(3)	5.86(9)	3.87(5)	173(6)	3.59(5)
cutoff z_c	7300(2700)	52.9(4.7)	49.0(4.3)	15.7(2.4)	9.4(1.3)	1200(300)	10.7(1.6)
exponent τ	2.5(1)	1.3(1)	0.91(10)	1.1(2)	1.1(2)	1.03(7)	1.3(2)
size of giant component	1193488	44337	14845	13861	5835	49002	6396
first initial only	892193	39709	12874	13324	5593	43089	6706
as a percentage	87.3(7)%	85.4(8)%	89.4(3)	84.6(8)%	71.4(8)%	88.7(1.1)%	57.2(1.9)%
2nd largest component	56	18	19	16	24	69	42
mean distance	4.4(2)	5.9(2)	4.66(7)	6.4(1)	6.91(6)	4.0(1)	9.7(4)
maximum distance	21	20	14	18	19	19	31
clustering coefficient C	0.072(8)	0.43(1)	0.414(6)	0.348(6)	0.327(2)	0.726(8)	0.496(6)

TABLE I. Summary of results of the analysis of seven scientific collaboration networks. Numbers in parentheses are standard errors on the least significant figures.

Распределение связности в метаболических сетях разных организмов



Свойства метаболических сетей

гистограмма длин биохимических путей E/Coli

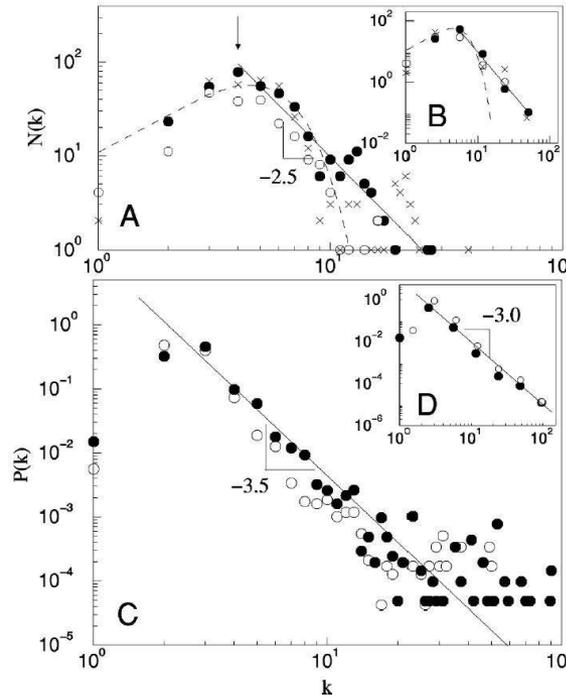
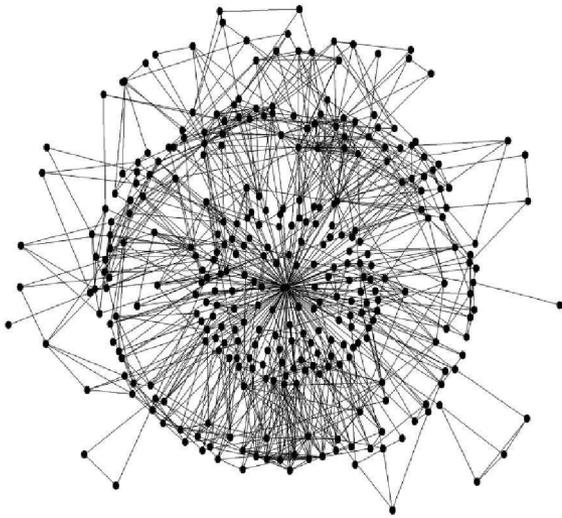
средняя длина пути (43 организма)

удельное среднее число выходящих и входящих связей

эффект удаления субстрата в сети E/Coli

Электрические устройства

Граф аналогового телевизора



Распределение связности
в малых устройствах (3
телевизора)

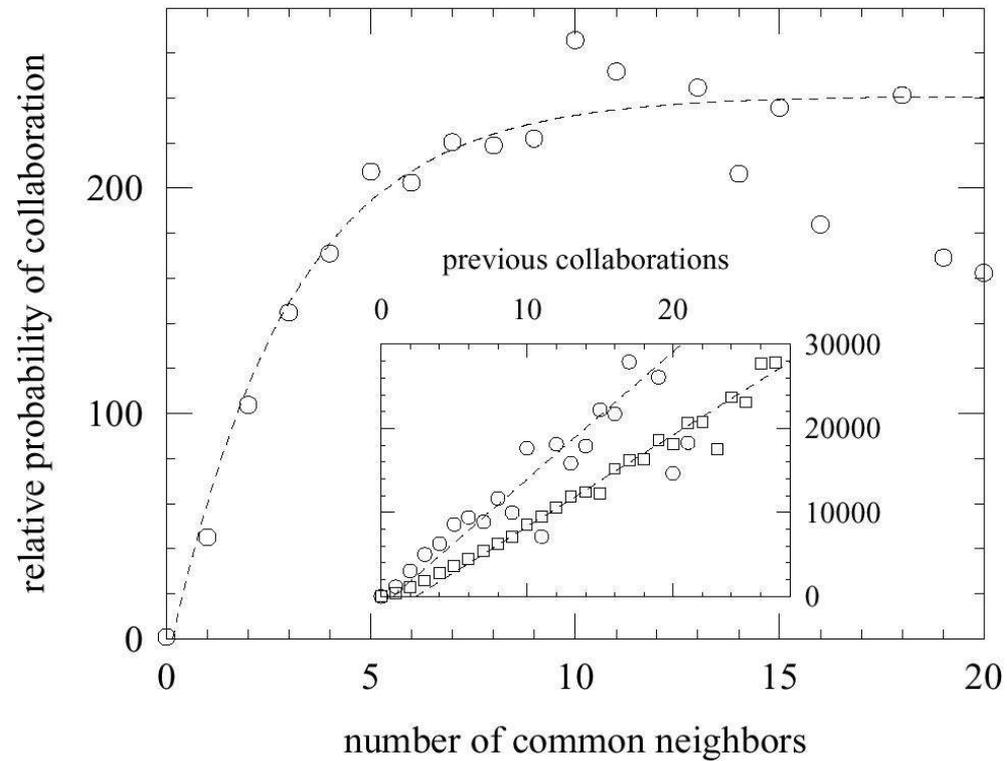
максимум при
связности=4, верхний
порог - 25-30

пунктир - пуассон

Распределение связности
в больших устройствах

верхний порог - 100

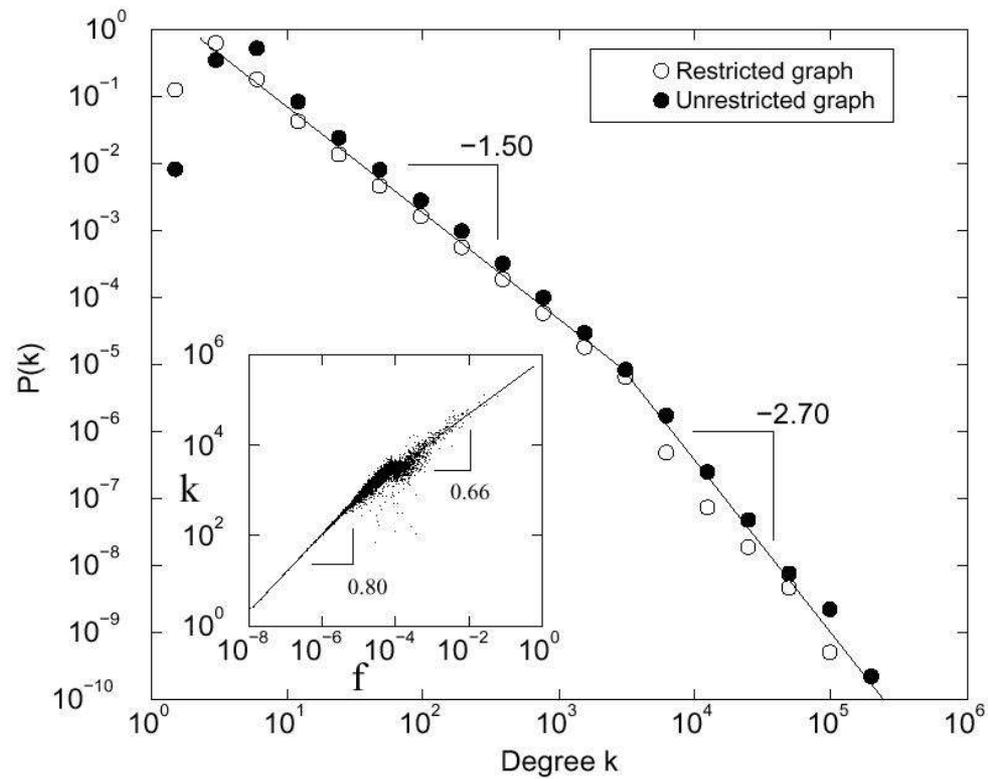
Вероятность сотрудничества в зависимости от числа общих предыдущих коллабораторов



Распределение связности в словарных сетях

малый мир (2-3)

безмасштабное рапределение связности



Свойства безразмерных сетей

- «малый мир» (короткие пути, свойство других случайных графов)
- кластеризация (много общих соседей)
- ускорение распространения сигнала
- синхронность
- параллелизм «вычислений»

структура



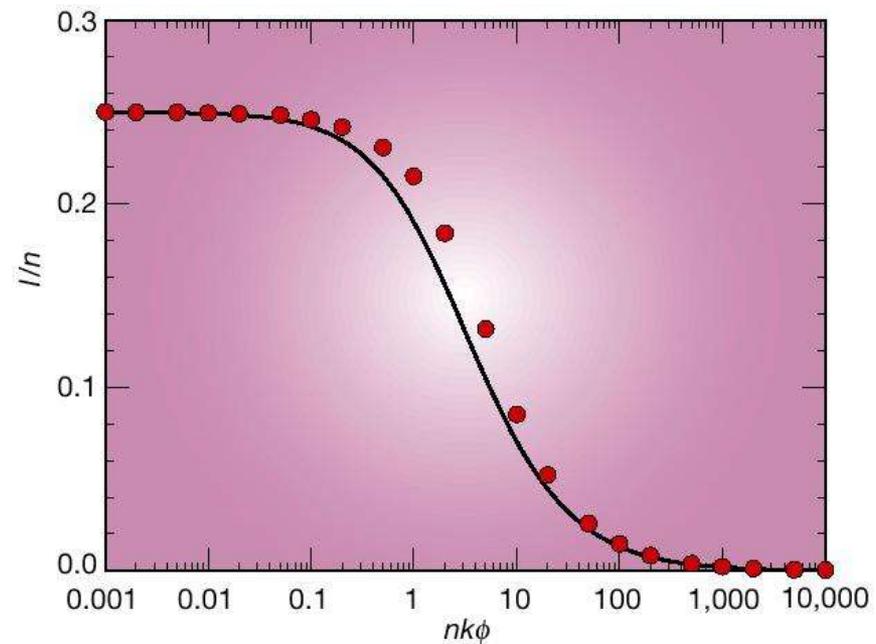
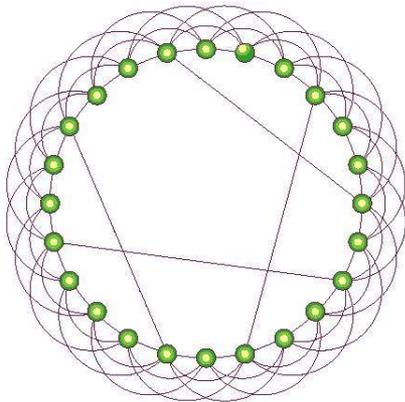
динамика

Легкость коммуникаций всей сети через каналы

+

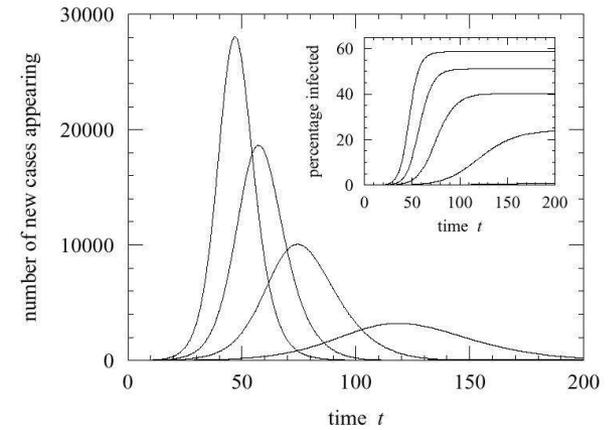
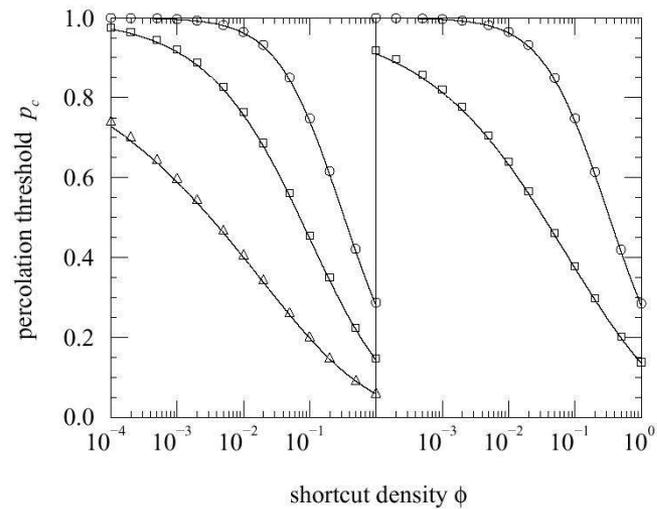
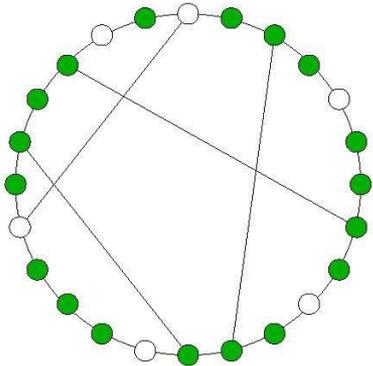
глобальная координация потоков информации

Аналитически решаемая модель «малого мира»



Удельная средняя длина пути в зависимости от среднего числа шунтов

Модель распространения эпидемий в «малом мире»



Растущие сети

Экспоненциальная сеть:

4) добавить узел

5) случайно соединить с другими

Безмасштабная сеть:

8) добавить узел

9) ?

Предпочтение в связывании

Popularity is attractive!

Следствие: The oldest is the richest

Структурные явления в растущих безразмерных сетях

Конденсация ребер

Корреляция между степенями соседей

Ускорение роста (www, Internet, citation)

Лексическое ядро (1000)

Сети: поведение при атаках и сбоях

Экспоненциальные сети повреждаются одинаково

Безмасштабные сети:

Устойчивы к сбоям и чувствительны к атакам (с увеличением интенсивности атаки - рост диаметра, исчезновение гигантской компоненты) – почему?

Эпидемии:

Перколяционный характер для *экспоненциальных сетей*

Нулевой порог для *безмасштабных сетей* с «толстыми хвостами», пандемии

Иммунизация ядра

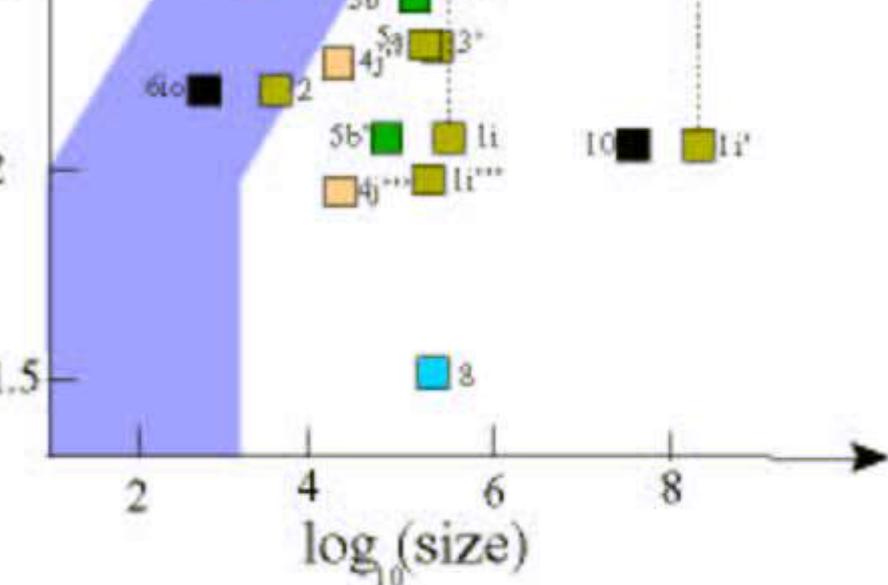
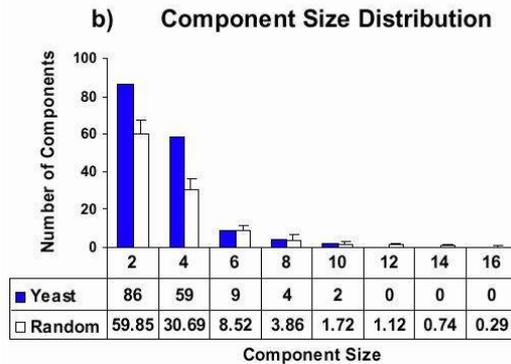
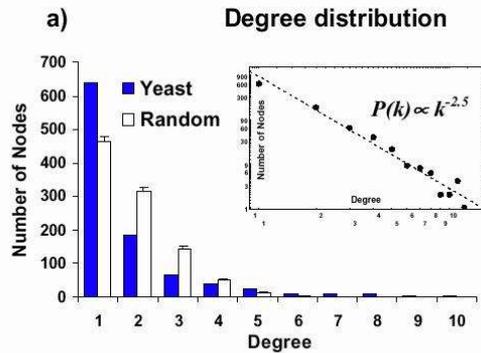
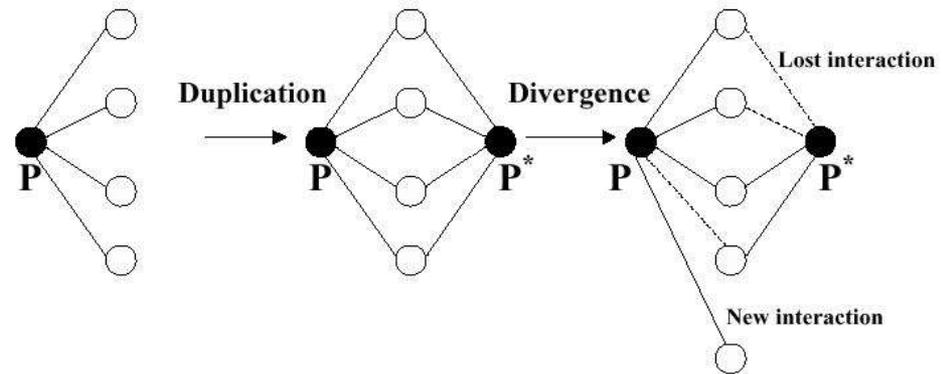


FIG. 24. Log-linear plot of the γ exponents of all the networks reported as having power-law (in-, out-) degree distributions (i.e., scale-free networks) vs. their sizes. The line $\gamma \sim 1 + \log_{10} t / 2.5$ is the estimate of the finite-size boundary for the observation of the power-law degree distributions for $\gamma > 2$. Here 2.5 is the range of degrees (orders) which we believe is necessary to observe a power law. The dashed line, $\gamma = 3$, is the resilience boundary (see Sec. XIV). This boundary is important for networks which must be stable to random breakdowns. The points are plotted using the data from Tab. I. Points for γ_0 from the same set of data are connected. The precision of the right points is about ± 0.1 (?) and is much worse for points in the grey region. There exists a chance that some of these nets are actually not in the class of scale-free networks.

Points: $1i$ and $1o$ are obtained from in- and out-degree distributions of the complete map of the nd.edu domain of the WWW; $1i'$ and $1o'$ are from in- and out-degree distributions of the pages of the WWW scanned by Altavista in October of 2000; $1o''$ is the γ_0 value from another fitting of the same data [61]; $1i'''$ is γ_i for domain level of the WWW in spring 1997 [117]; 2 is γ for the inter-domain level of the Internet in December 1998 [5]; $2'$ is γ for the network of operating systems one of days in December 1998 [96]; 3 is γ for the router level of the Internet in 1995 [5]; $3'$ is γ for the router level of the Internet in 2000 [104]; $4i$ is γ_i for citations of the ISI database 1981 - June 1997 [27]; $4i'$ is the result of the different fitting of the same data [99]; $4i'''$ is another estimate obtained from the same data [94,95]; $4j$ is γ_i for citations of the Phys. Rev. D 11-50 (1975-1994) [27]; $4j'$ is the different fitting of the same data [99]; $4j''$ is another estimate from the same data [94,95]; $4j'''$ is γ_i for citations of the Phys. Rev. D (1982-June 1997) [100]; $5a$ is the γ exponent for the collaboration network of movie actors [55]; $5a'$ is the result of another fitting for the same data [102]; $5b$ is γ for the collaboration network of MEDLINE [18]; $5b'$ is γ for the collaboration net collected from mathematical journals [15]; $5b''$ is γ for the collaboration network collected from neuro-science journals [15]; $6i_0$ is $\gamma_i = \gamma_o$ for networks of metabolic reactions [41]; 7 is γ of the network



Свойства безразмерных сетей обусловлены их эволюционной природой; или селекция?



Сети белковых контактов в дрожжах и в случайных графах

а - число соседей

б - размеры
КОМПОНЕНТ

Эффект дупликации генов на их продукты, что взаимодействуют с белками

Свойства безразмерных сетей

- «малый мир» (короткие пути)
- кластеризация (общие соседи)
- ускорение распространения сигнала
- синхронность
- параллелизм «вычислений»

структура



динамика

Легкость коммуникаций всей сети через каналы

+

глобальная координация потоков информации