



# Структура функциональных сетей

*к.ф.м.-н. И.И. Титов*

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia



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



**Следующий этап постгеномной эры - исследование взаимодействий**

- между генами**
- между их продуктами**

Индуктивные и репрессивные взаимодействия



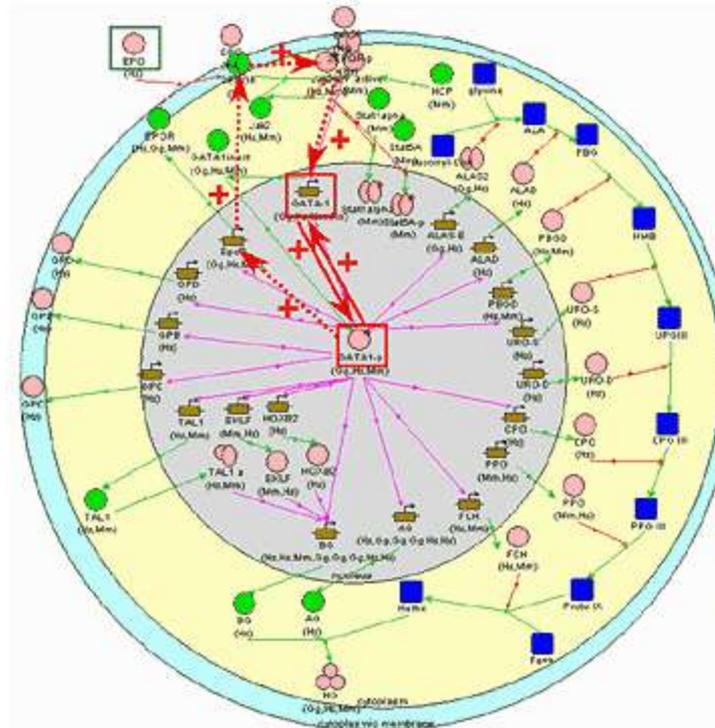
# Положительная обратная связь



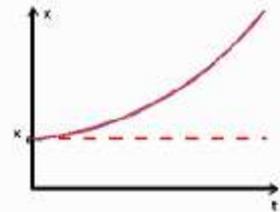
Положительные обратные связи обеспечивают эффективное отклонение контролируемого параметра  $X$  от его текущего значения. Они играют ключевую роль в генных сетях роста и дифференцировки клеток, морфогенеза органов, роста и развития организмов, при которых система непрерывно уходит от своего текущего состояния. Например, генная сеть дифференцировки и созревания эритроцитов контролируется по механизму положительной обратной связи. Исследование и моделирование генных сетей, контролирующих функции кроветворения, имеют фундаментальное значение для понимания процессов клеточной дифференцировки, а в прикладном плане - для поиска путей фармакологической коррекции патологий системы кроветворения.

БАЗА ДАННЫХ GENENET:

ФРАГМЕНТ ГЕННОЙ СЕТИ ДИФФЕРЕНЦИРОВКИ И СОЗРЕВАНИЯ ЭРИТРОЦИТОВ



Принципиальная схема регуляторного контура с отрицательной обратной связью





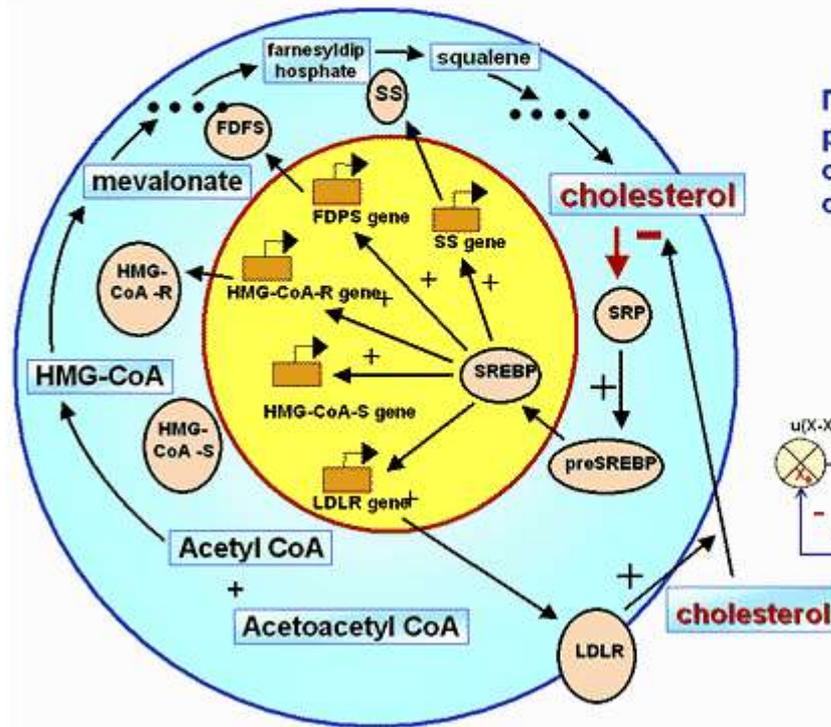
# Отрицательная обратная связь



Ключевую роль в функционировании генных сетей, обеспечивающих гомеостаз, играют регуляторные контуры с отрицательными обратными связями, поддерживающие контролируемый параметр вблизи определенного уровня. На рисунке показан центральный фрагмент генной сети, обеспечивающей биосинтез холестерина и постоянство его концентрации в клетках. Обозначения таковы. Центральный круг - ядро, периферия - цитоплазма, внешняя полоса - мембрана клетки; гены - прямоугольники, белки - круги, метаболиты - синие квадраты; реакции и регуляторные события - стрелки. Путь биосинтеза холестерина из ацетил-коэнзима А контролируется по механизму отрицательной обратной связи. Центральный регулятор генов, кодирующих ферменты этого пути, транскрипционный фактор SREBP, активирующий транскрипцию каскада этих генов и тем самым усиливающий продукцию холестерина. Фактор SREBP образуется из предшественника preSREBP под действием стерол-зависимой протеазы. При повышении уровня холестерина активность протеазы подавляется, что снижает скорость образования фактора SREBP и его концентрацию. Тем самым снижается активность генов, кодирующих ферменты этого пути, и уровень холестерина нормализуется. Так работает отрицательная обратная связь, контролирующая концентрацию холестерина. Всего эта генная сеть включает несколько сотен генов.

БАЗА ДАННЫХ GENENET:

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

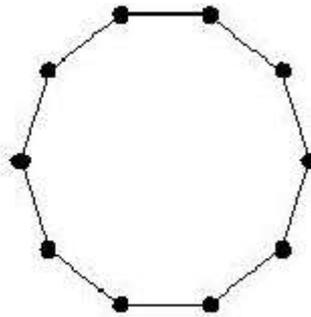


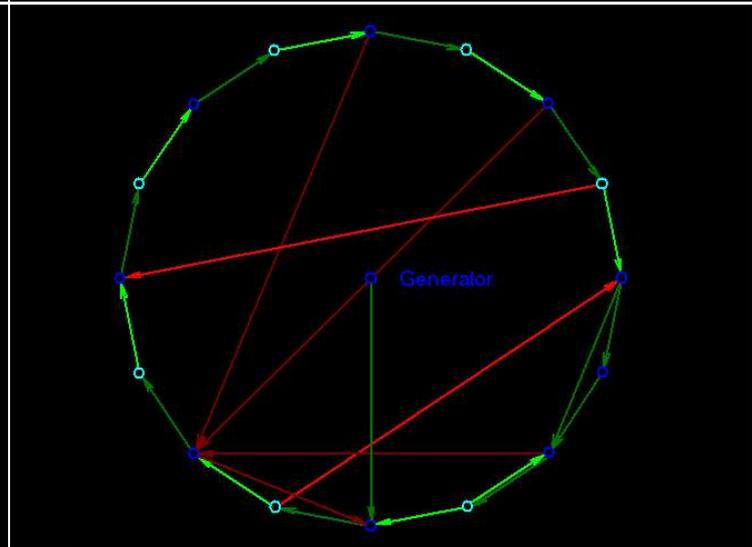
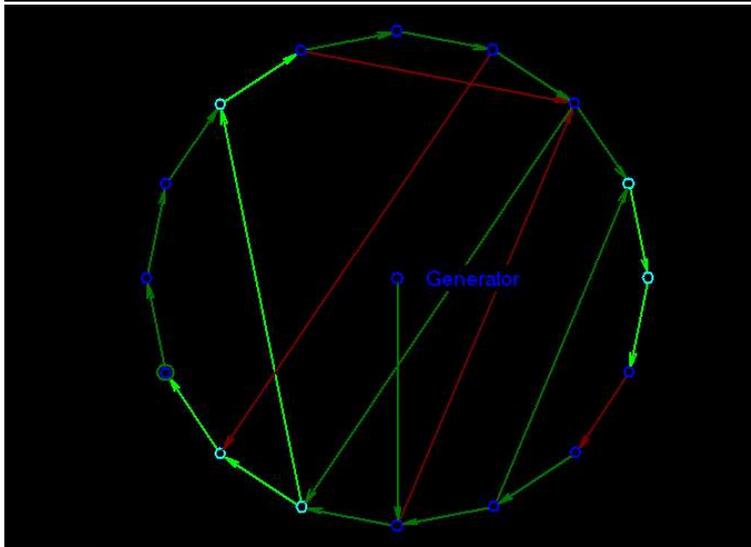
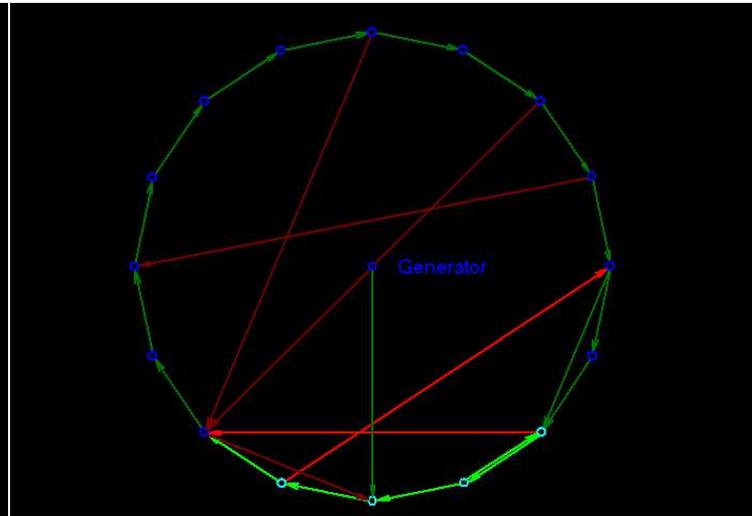
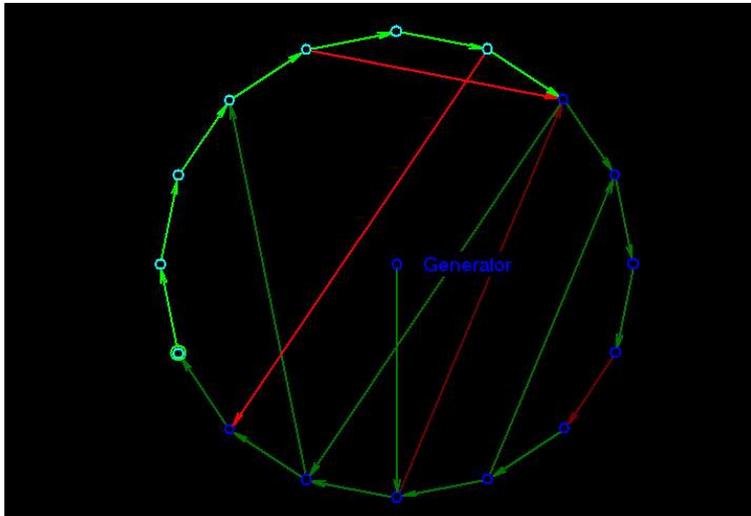
Принципиальная схема регуляторного контура с отрицательной обратной связью





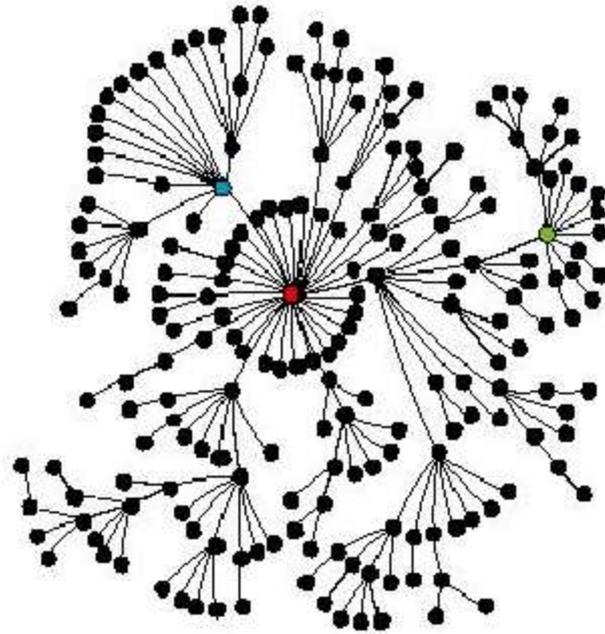
# Циклическая и последовательная активация

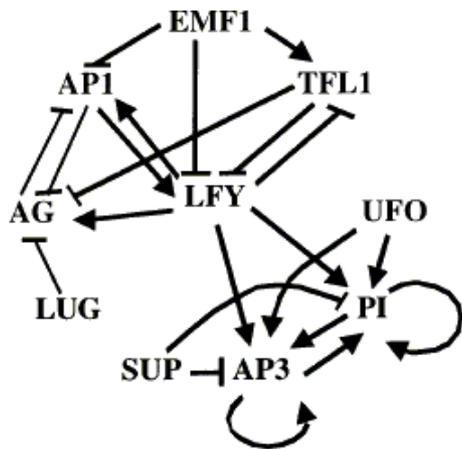






# Кассетная активация

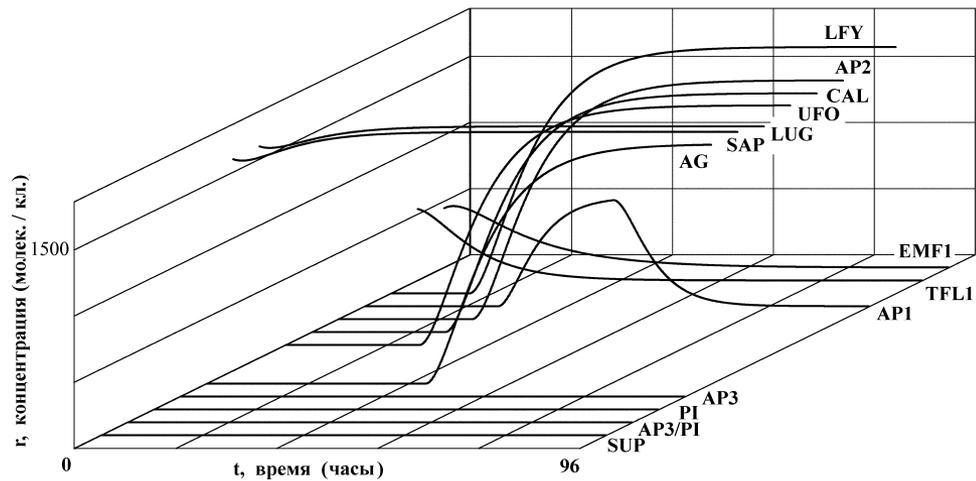
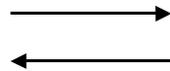
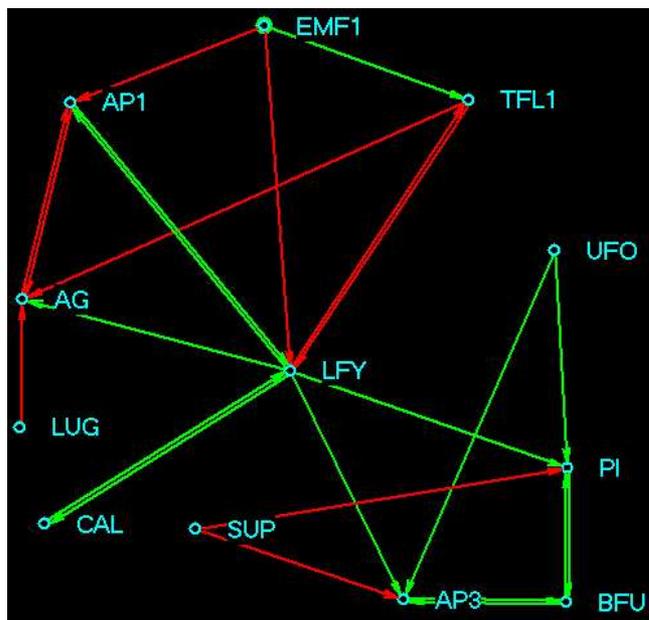




## 1. Прямая задача для генных сетей

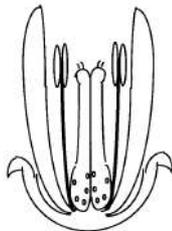
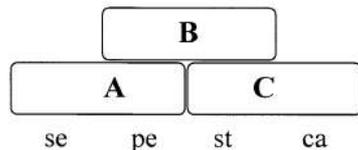
Дана структура и параметры - найти поведение

## 2. Обратная задача найти структуру и параметры

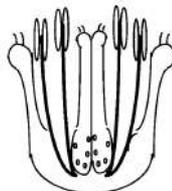
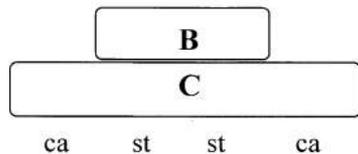




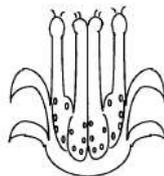
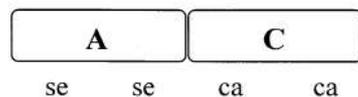
wild type



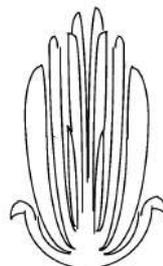
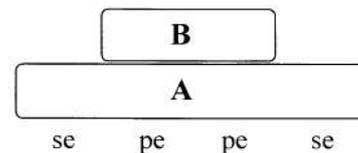
*ap1*  
Mutants  
(A-)



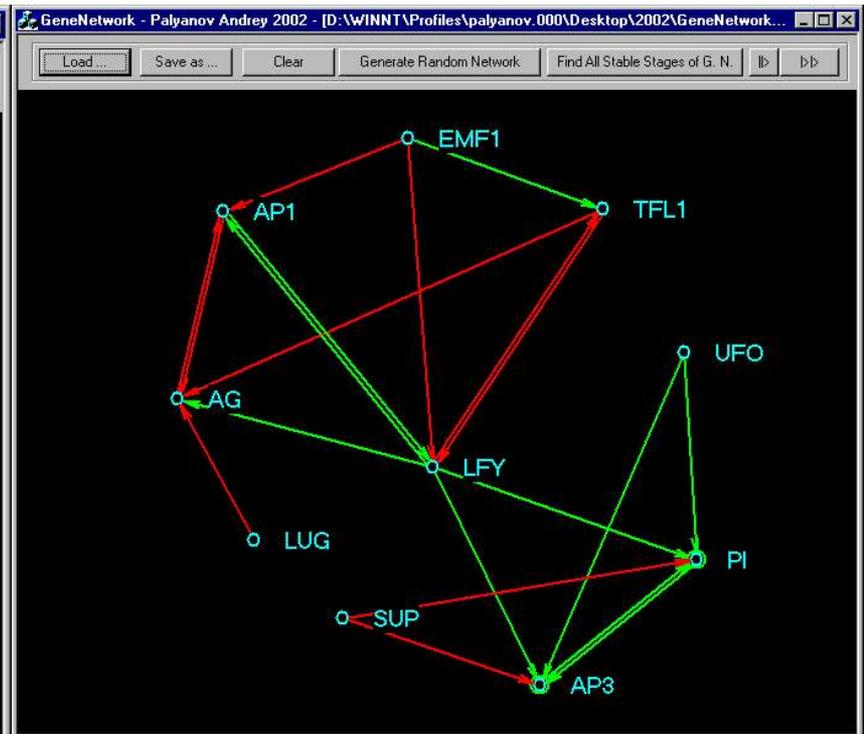
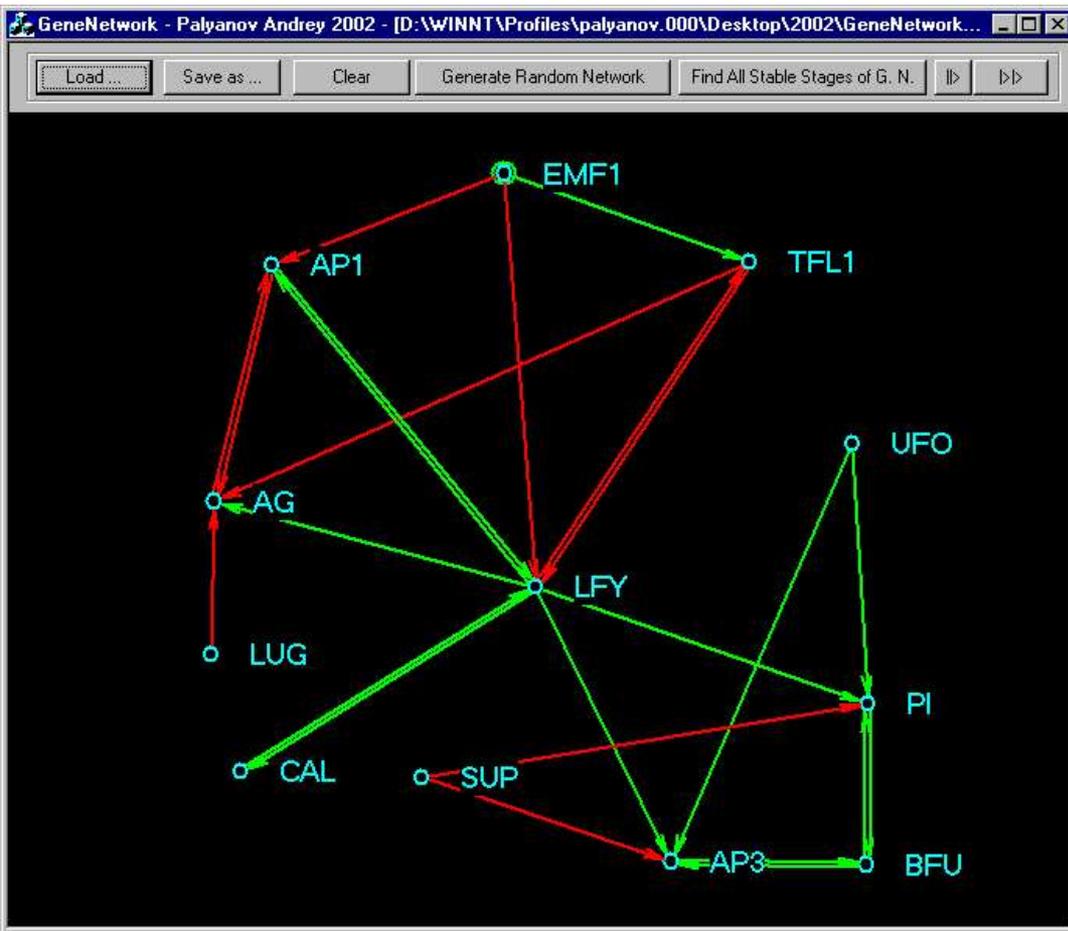
*ap3* or *pi*  
Mutants  
(B-)

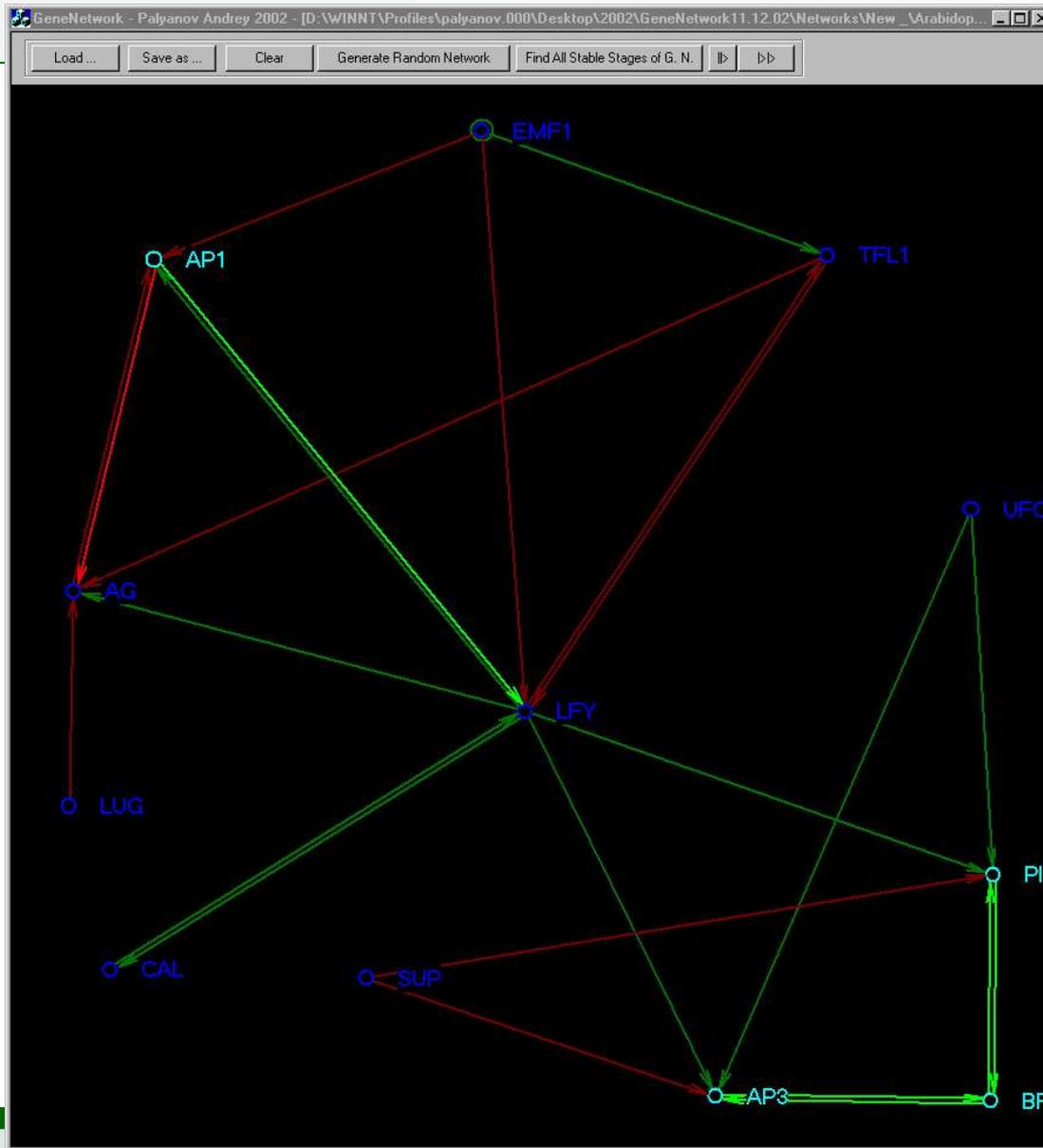


*ag*  
Mutants  
(C-)



- $[00] + [0010] \rightarrow [0001000000]$  (A = sepals)
- $[11] + [0010] \rightarrow [0001000110]$  (AB = petals)
- $[11] + [0001] \rightarrow [0000100110]$  (BC = stamen)
- $[00] + [0001] \rightarrow [0000100000]$  (C = carpels)
- $[00] + [1000] \rightarrow [1100000000]$  (flower inhibition)
- $[11] + [1000] \rightarrow [1100000110]$  (6th attractor)





Поиск устойчивых состояний генной сети - А. Пальянов ...

N	E...	T...	L...	A...	C...	L...	U...	B...	A...	A...	PI	S...
1	0	0	0	0	0	0	0	0	1	0	0	0
2	0	0	0	0	0	0	0	1	1	1	1	0
3	0	0	0	1	0	0	0	0	0	0	0	0
4	0	0	0	1	0	0	0	1	0	1	1	0
5	1	1	0	0	0	0	0	0	0	0	0	0
6	1	1	0	0	0	0	0	1	0	1	1	0



# Безразмерные сети



- пищевые и экологические
- метаболические
- генные
- лингвистические
- социальные (соавторов, содиректоров, актеров и пр)
- интернет
- радиоэлектронные и электрические
- программистские (большие программы)



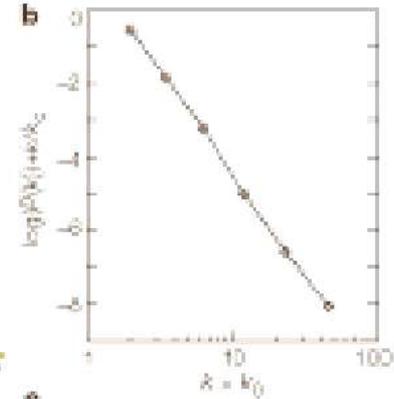
# Протеома E.Coli: 1870 белков и 2240 идентифицированных взаимодействий



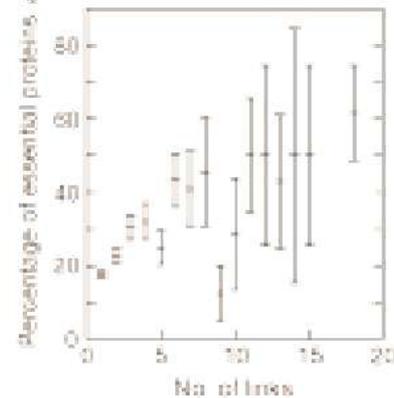
a



b



c





## Свойства безразмерных сетей



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

структура



динамика

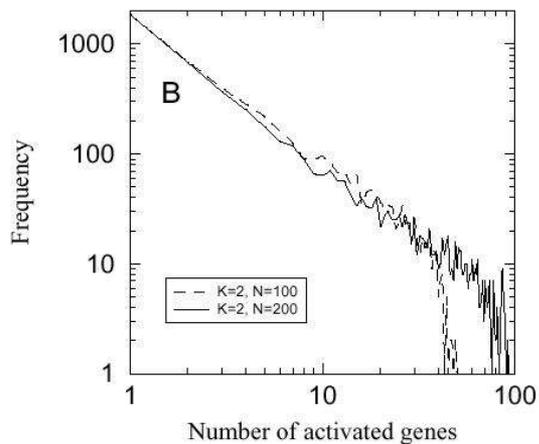
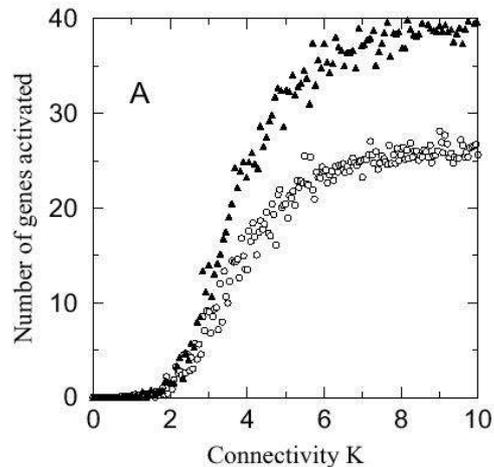
Легкость коммуникаций всей сети через каналы

+

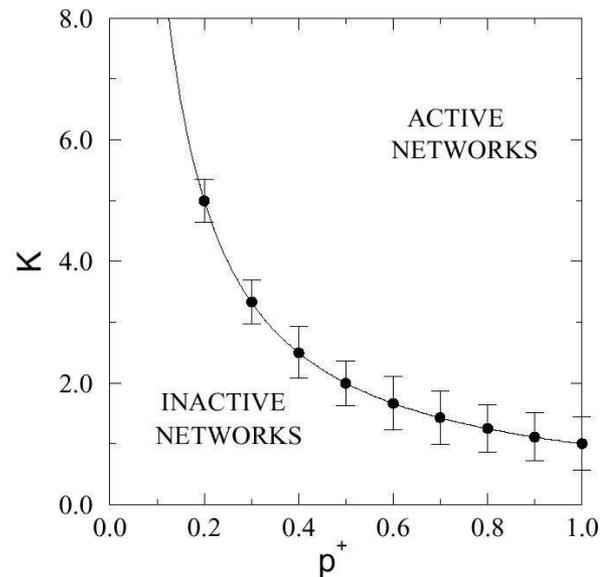
глобальная координация потоков информации



# Первые попытки анализа взаимодействий генов - модель случайных графов Эрдеша и Реньи



**Фазовый переход  
в случайной генной сети**

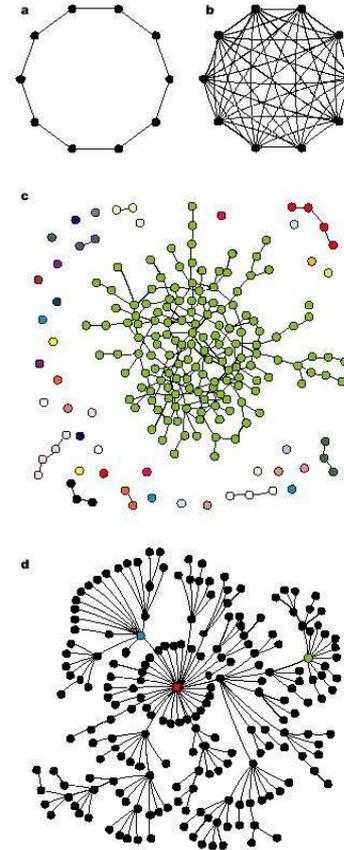
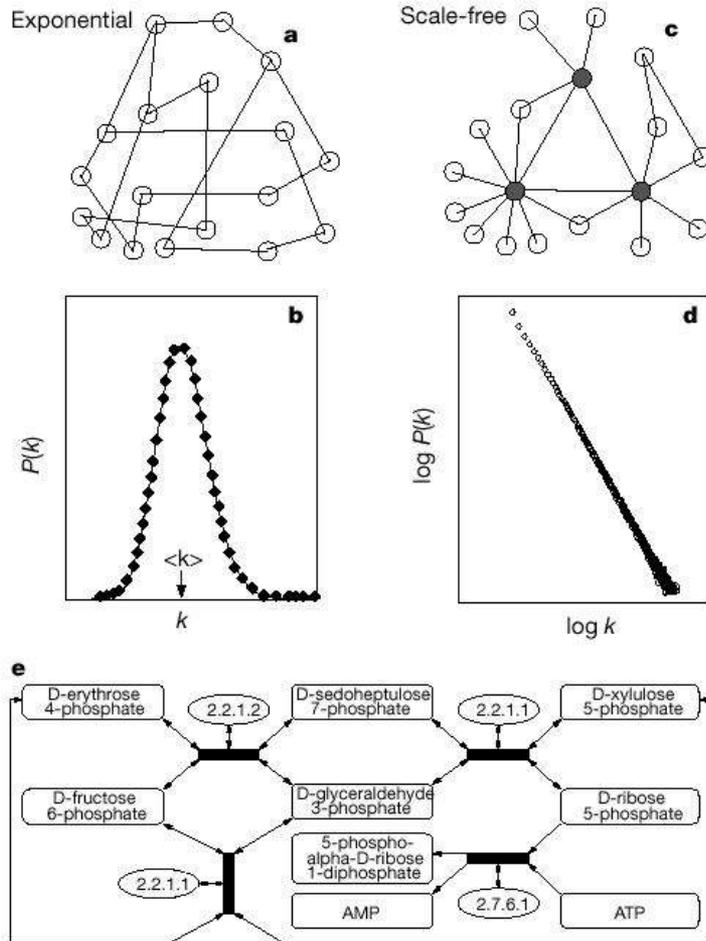


**Фазовая диаграмма случайной  
генной сети**

**$p$  - вероятность активации**



# Что такое безразмерная сеть



Архитектура регулярных  
и случайных сетей



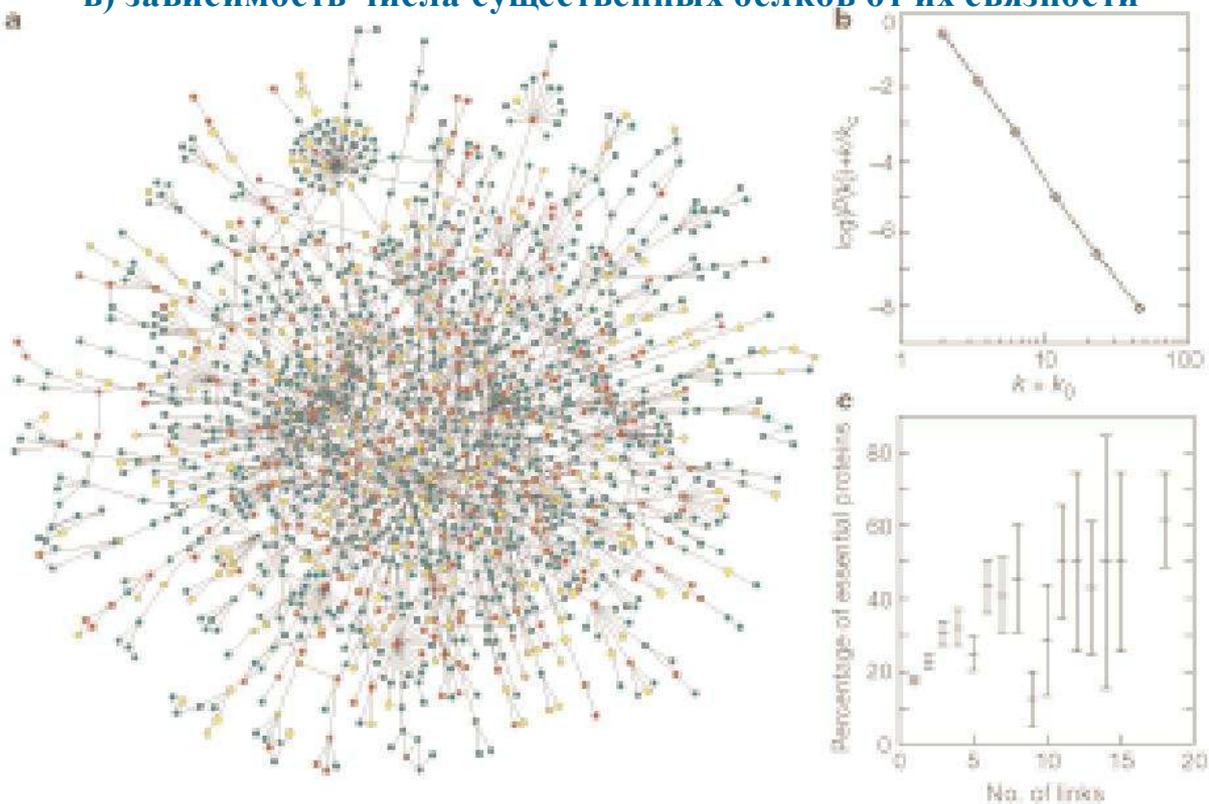
# Характеризация протеомы E.Coli

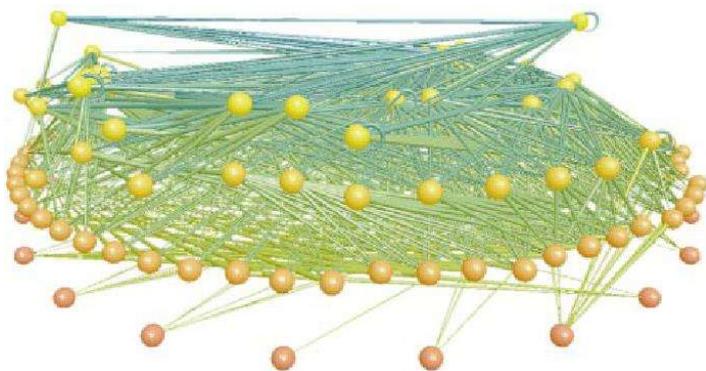


а) карта бело-белковых взаимодействий; наибольший кластер содержит 78% белков

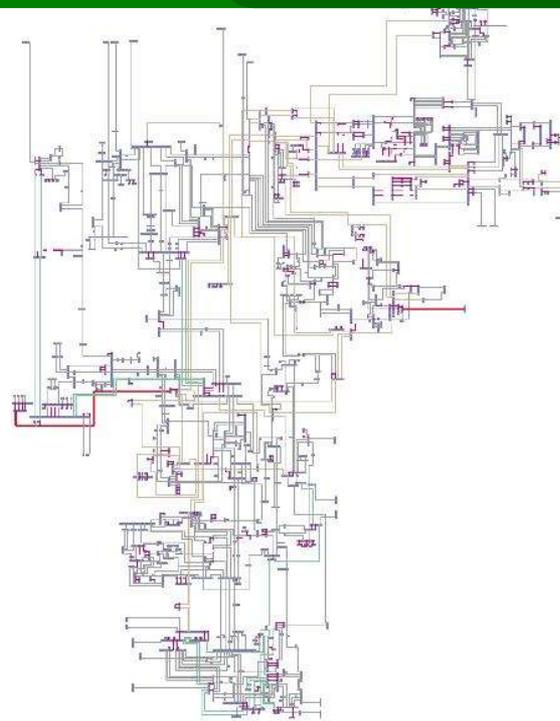
б) распределение связности

в) зависимость числа существенных белков от их связности

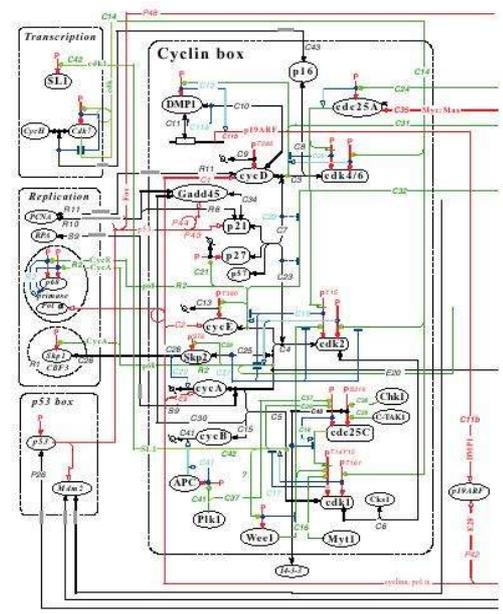




Пищевая карта озера  
Литл Рок



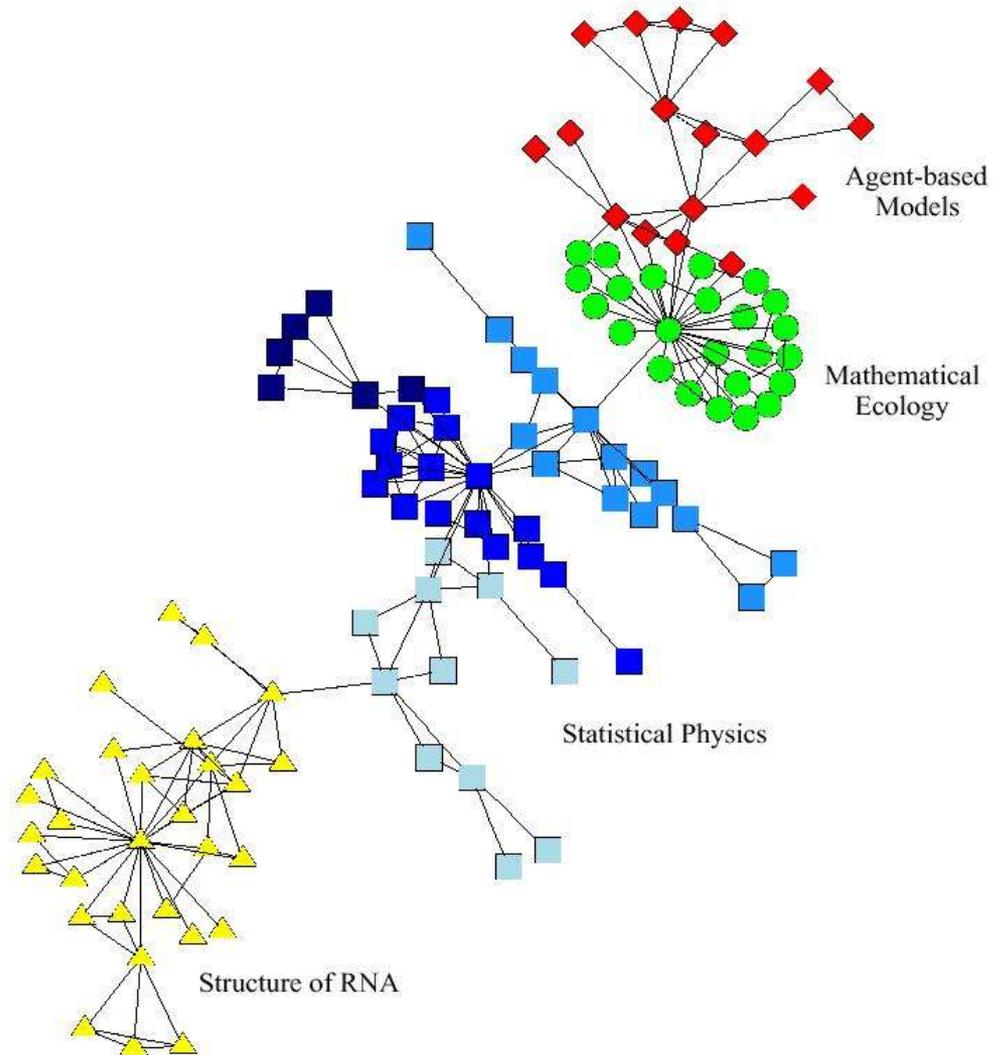
Электросеть штата Нью Йорк



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



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

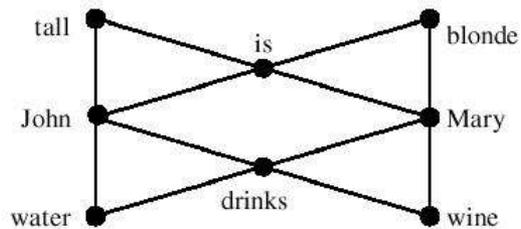




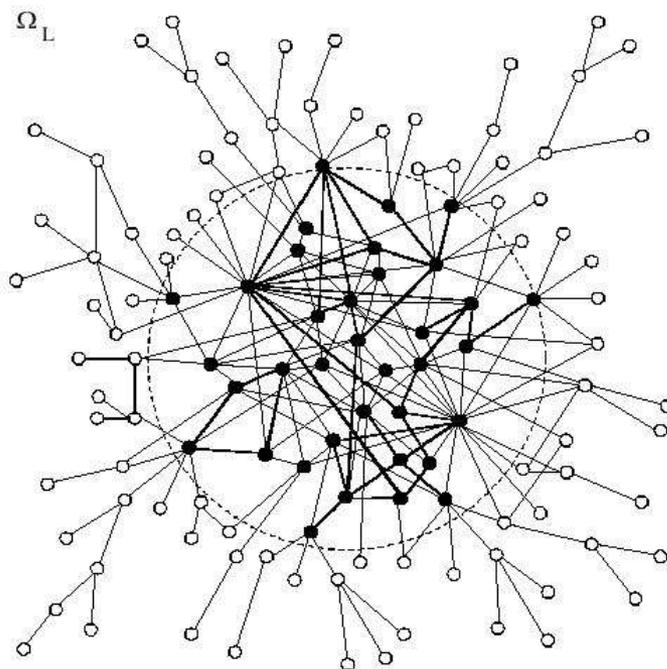
# Словарные сети



(a)



(b)





## Параметры реальных сетей



Network	$N$	$\ell$	$C$	$C_{\text{rand}}$
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



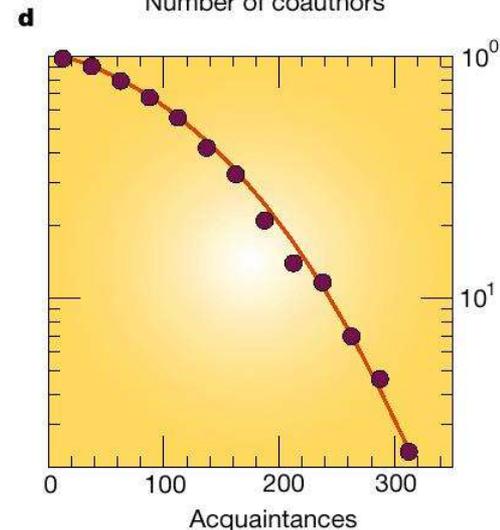
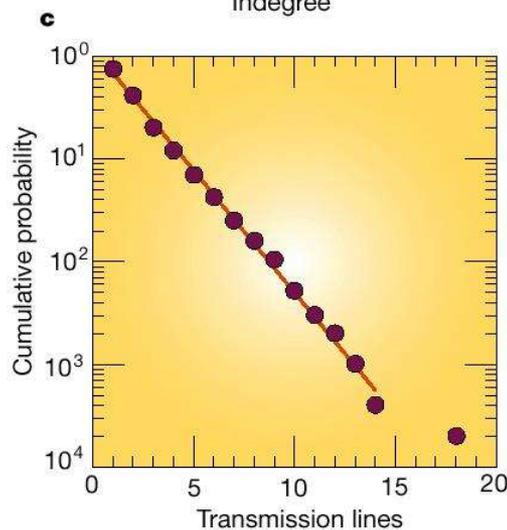
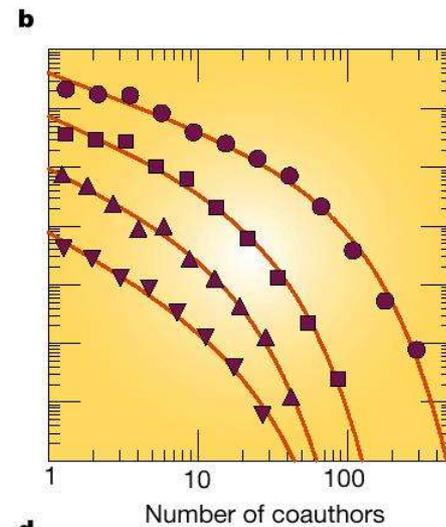
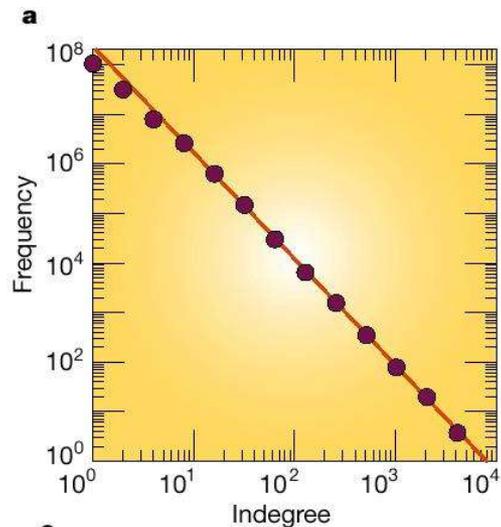
# Распределение связности реальных сетей

интернет

соавторство

электросеть США и Канады

социальная сеть мормонов





# Циклы в метаболических сетях

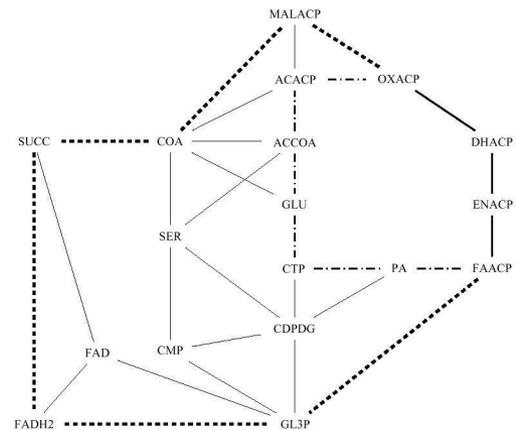
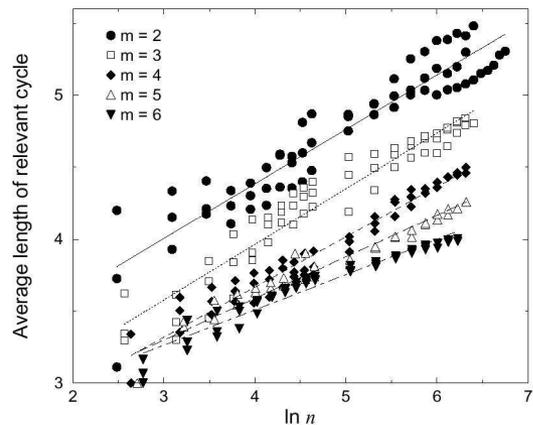


TABLE I: Cycle Structure of Metabolic Networks.

Model	$ C $	3	4	5	6	7	8	9	$\Sigma$
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

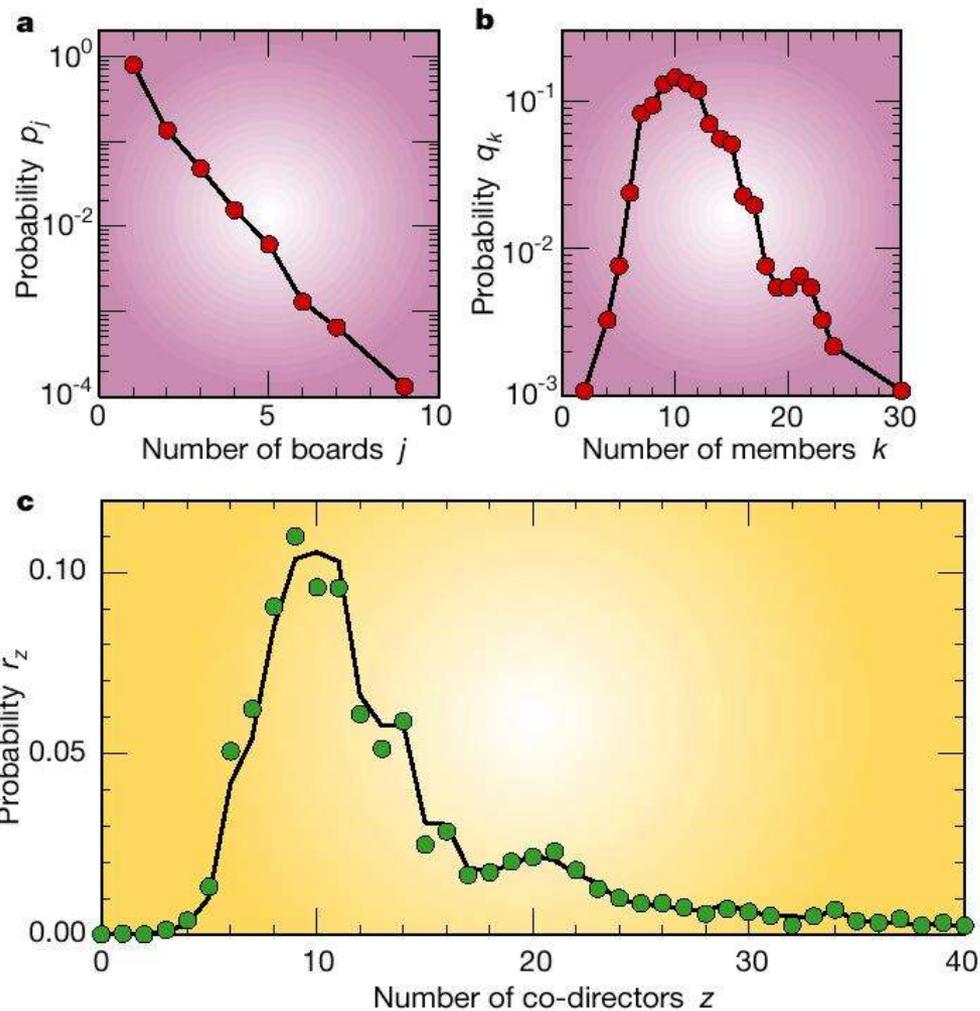


## Сеть директоров корпораций

распределение числа постов на  
одного директора

распределение корпораций по  
числу директоров

распределение по числу со-  
директоров





## Характеристики кластеризации в трех сетях

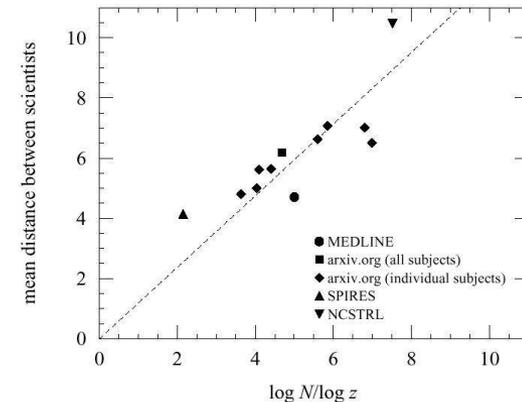
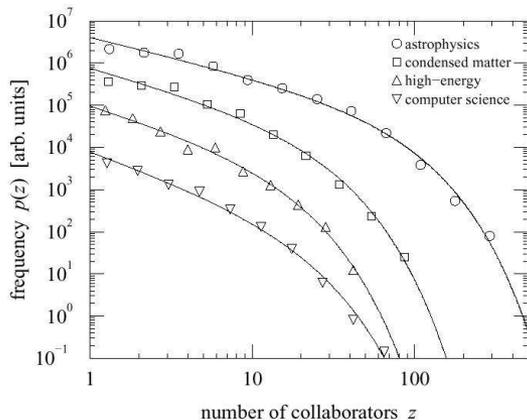


Network	Clustering $C$	
	Theory	Actual
Company directors	0.590	0.588
Movie actors	0.084	0.199
Biomedical authors	0.042	0.088

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](http://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

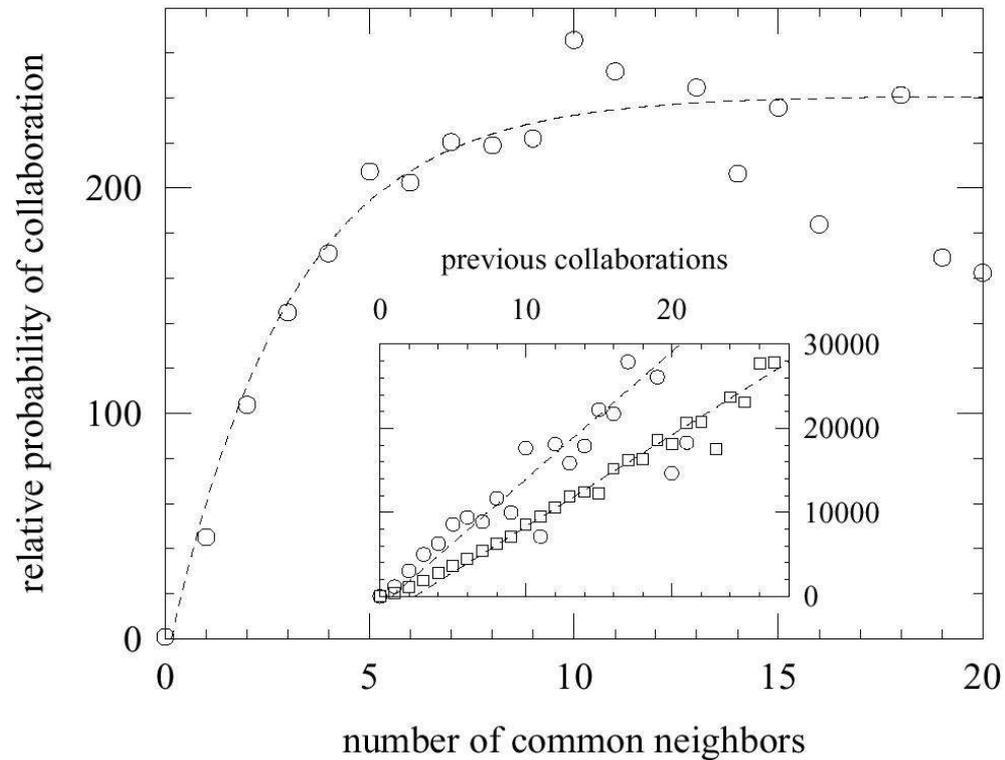


	MEDLINE	Los Alamos e-Print Archive				SPIRES	NCSTRL
		complete	astro-ph	cond-mat	hep-th		
total papers	2156769	98502	22029	22016	19085	66652	13169
total authors	1388989	52909	16706	16726	8361	56627	11994
first initial only	1006412	45685	14303	15451	7676	47445	10998
mean papers per author	5.5(4)	5.1(2)	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 $\tau$	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.



# Вероятность сотрудничества в зависимости от числа общих предыдущих коллабораторов



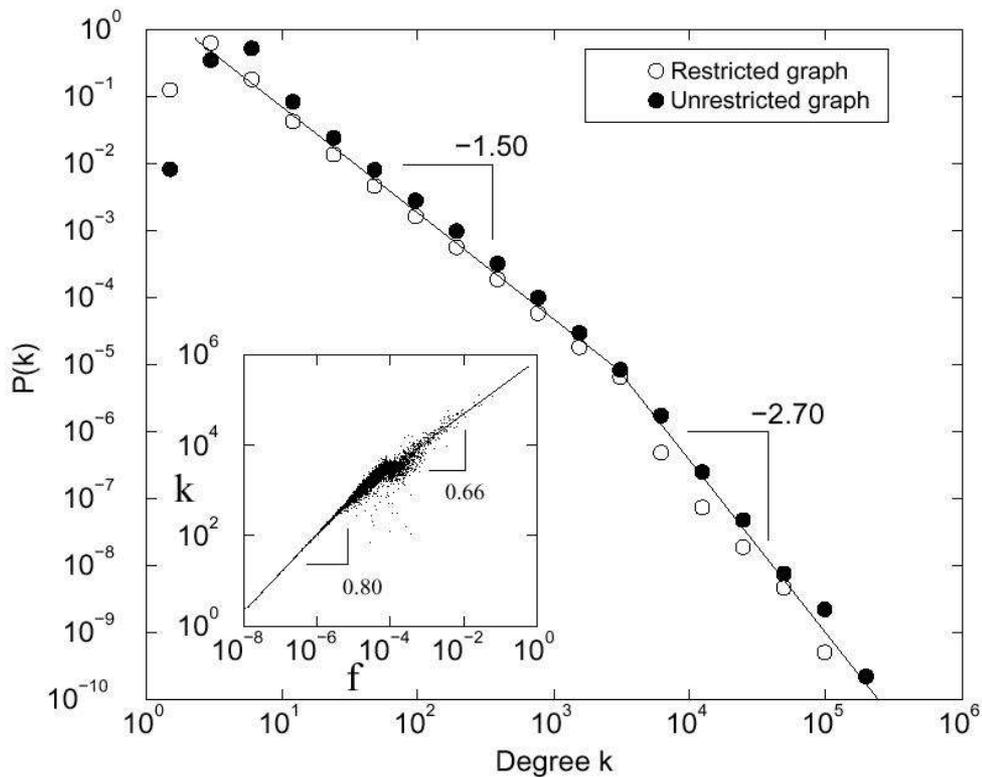


# Распределение связности в словарных сетях



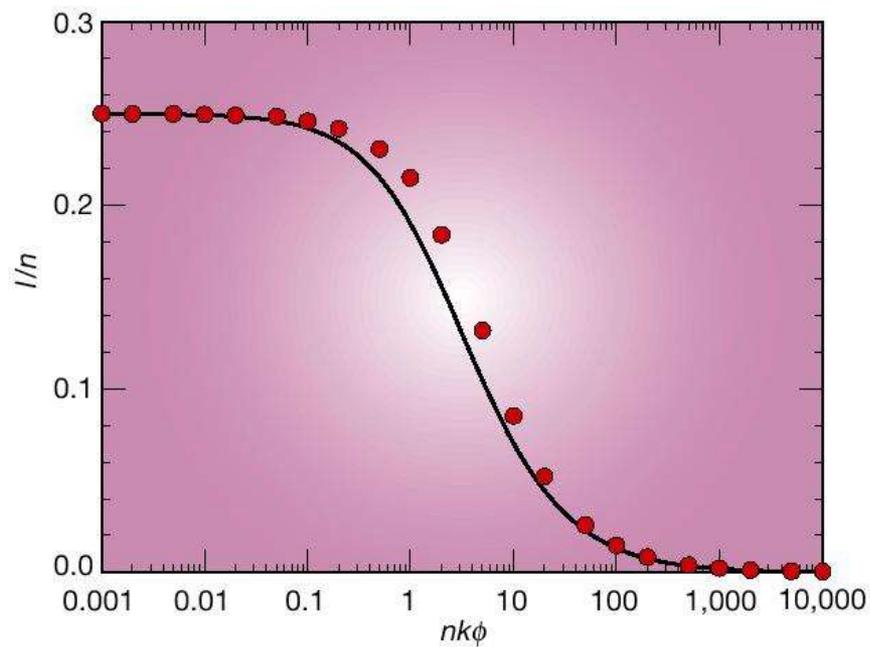
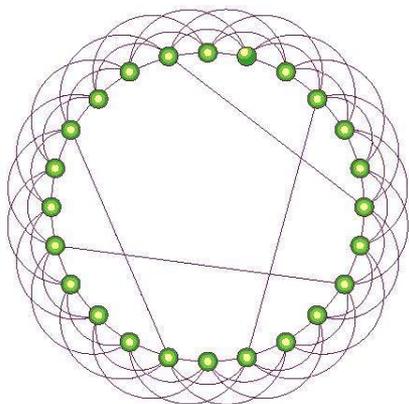
малый мир (2-3)

безмасштабное рапределение связности





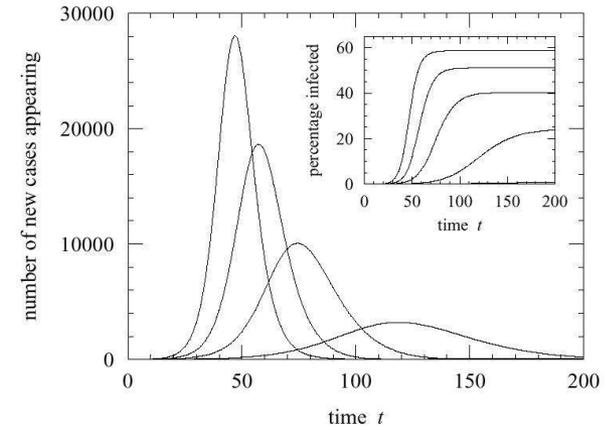
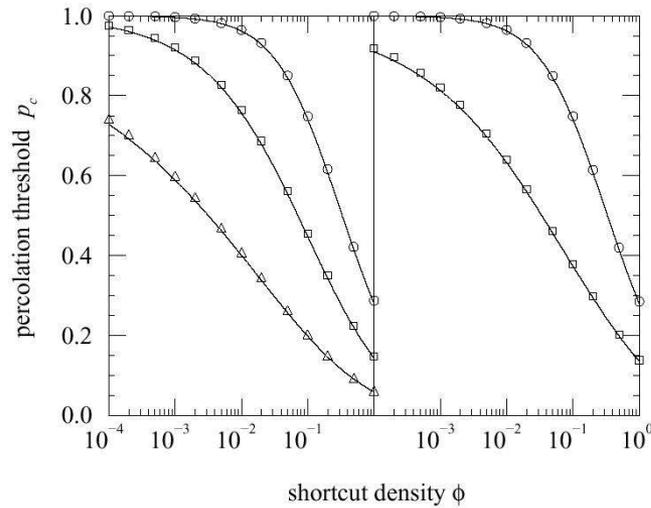
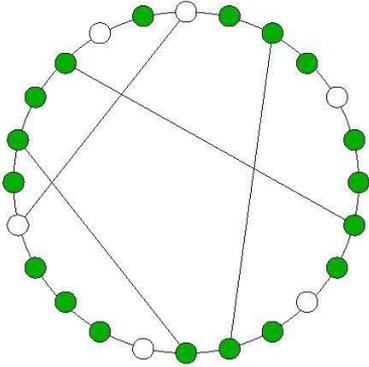
# Аналитически решаемая модель «малого мира»



Удельная средняя длина пути в зависимости от среднего числа шунтов

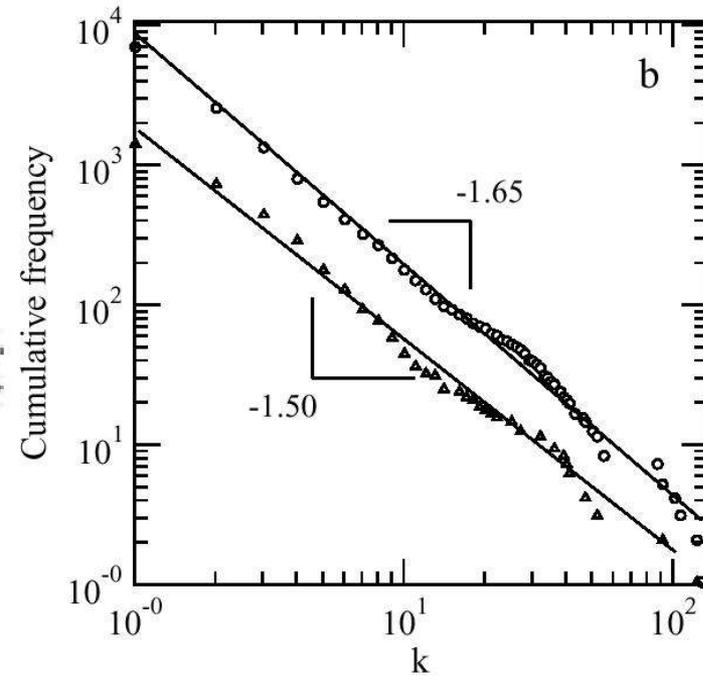
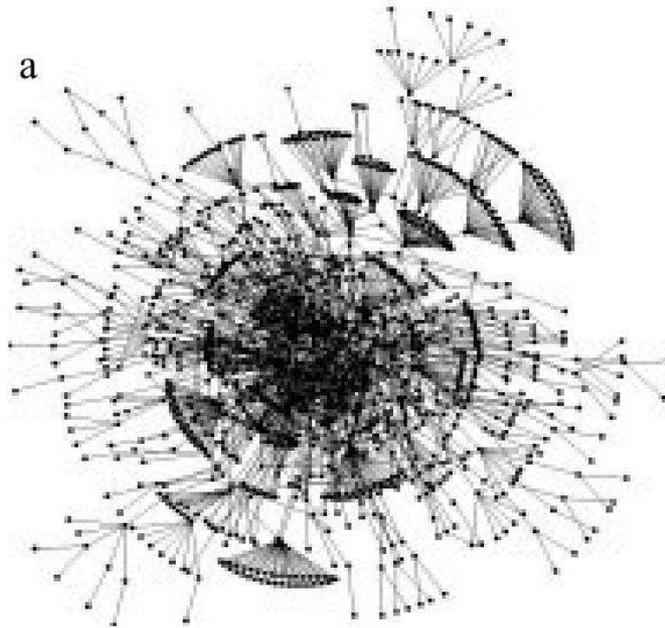


# Модель распространения эпидемий в «малом мире»



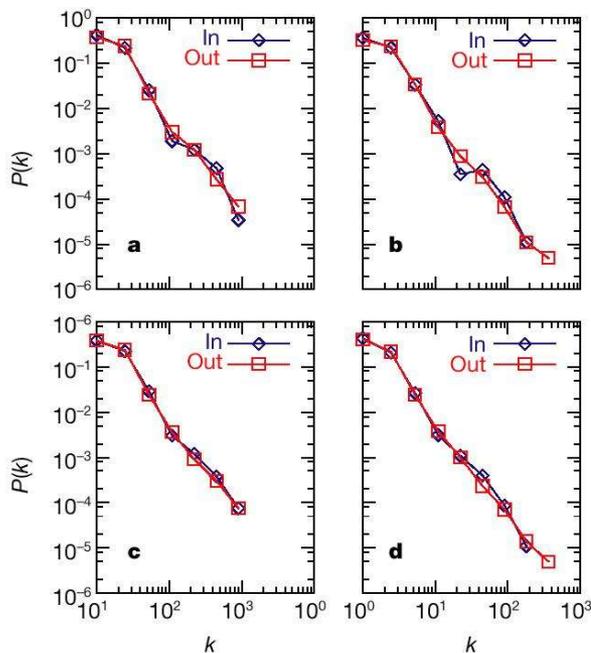


# Диаграмма классов Java Development Framework 1.2





## Распределение связности в метаболических сетях разных организмов



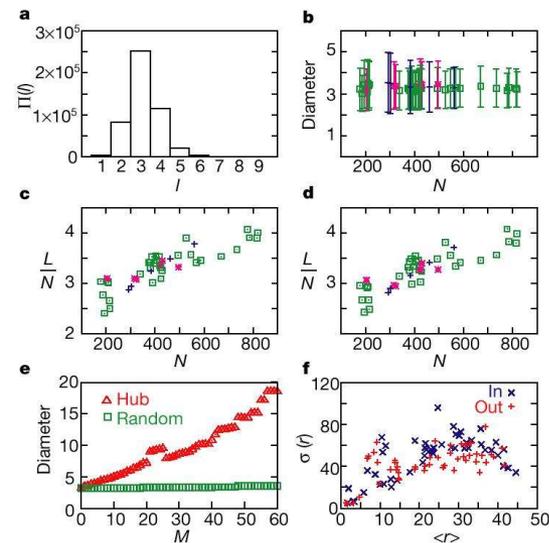
## Свойства метаболических сетей

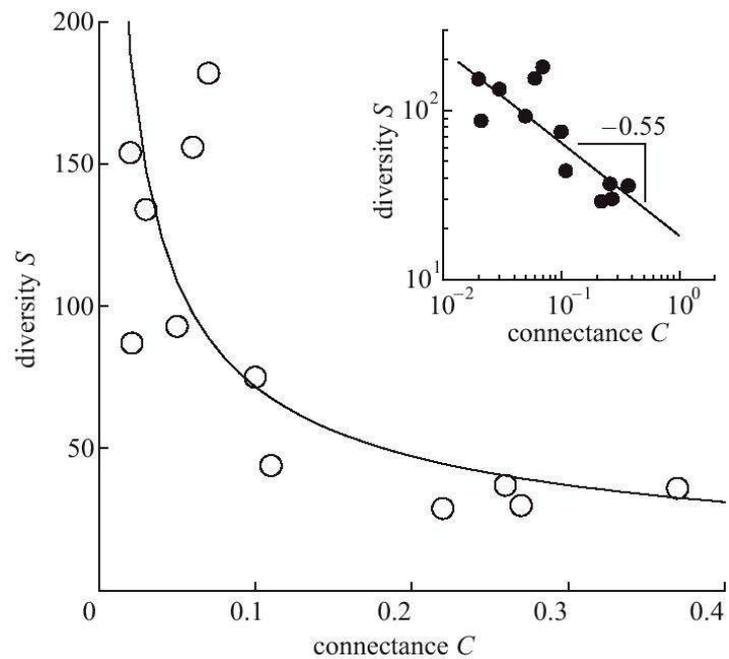
гистограмма длин биохимических путей E/Coli

средняя длина пути (43 организма)

удельное среднее число выходящих и входящих связей

эффект удаления субстрата в сети E/Coli

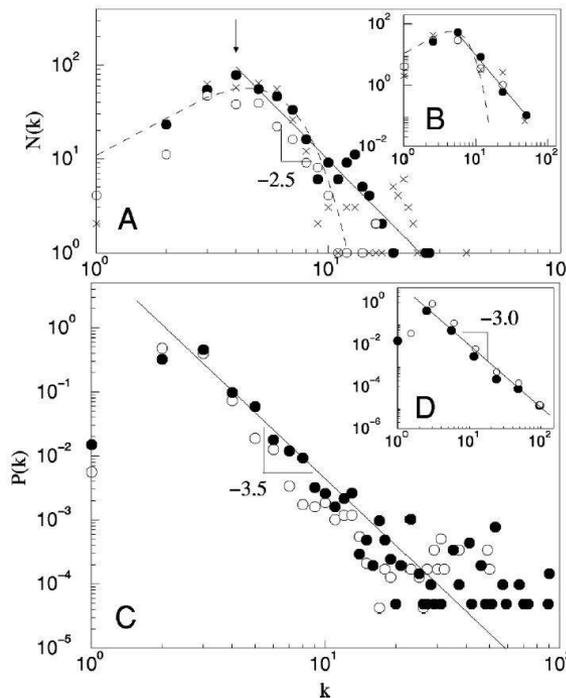
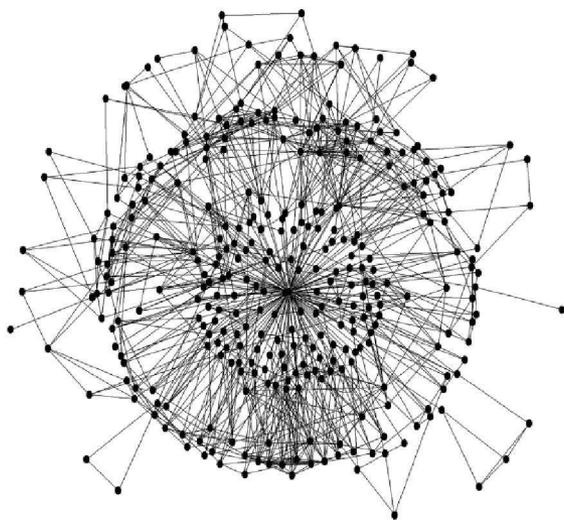




**ЧИСЛО ВИДОВ В ЗАВИСИМОСТИ ОТ СВЯЗНОСТИ  
В 12 ЭКОЛОГИЧЕСКИХ СООБЩЕСТВАХ**



## Граф аналогового телевизора



Распределение  
связности в малых  
устройствах  
(3 телевизора)

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

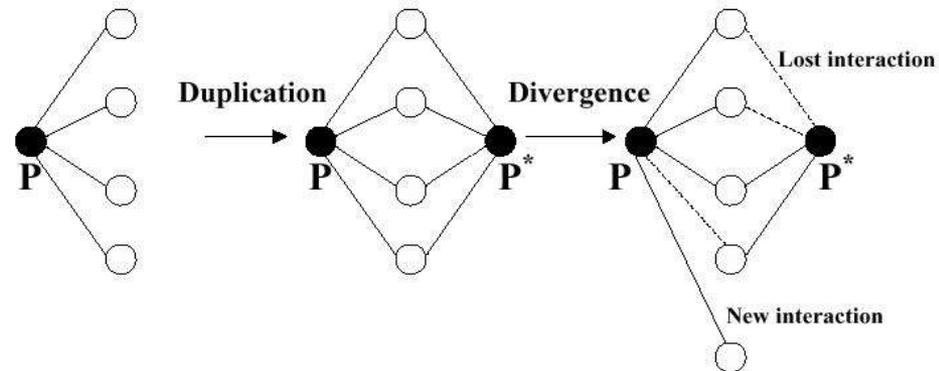
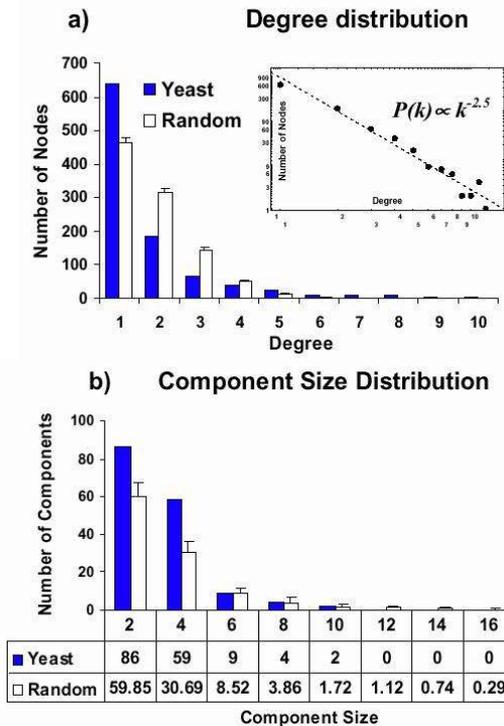
пунктир - пуассон

Распределение  
связности в больших  
устройствах

верхний порог - 100



# Свойства безразмерных сетей обусловлены их эволюционной природой



Сети белковых контактов в дрожжах и в случайных графах

а - число соседей

б - размеры компонент

Эффект дубликации 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



## Возможные постановки задач



- **идентификация отклонений**  
(анализ частных случаев)
- **исследование динамики сетей и оценка их «качества»**
- **восстановление эволюционных требований** (обратная задача: какова структура сети, оптимизирующей ту или иную функцию, проекция на природные сети)
- **анализ сетей других комбинаторных объектов** (РНК, белки)