

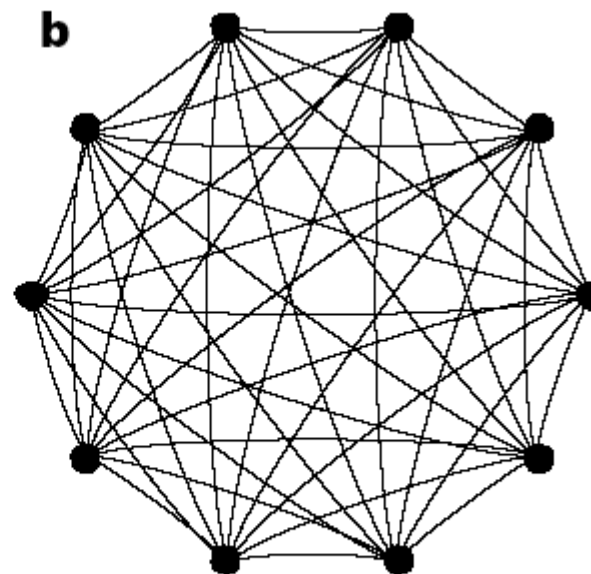
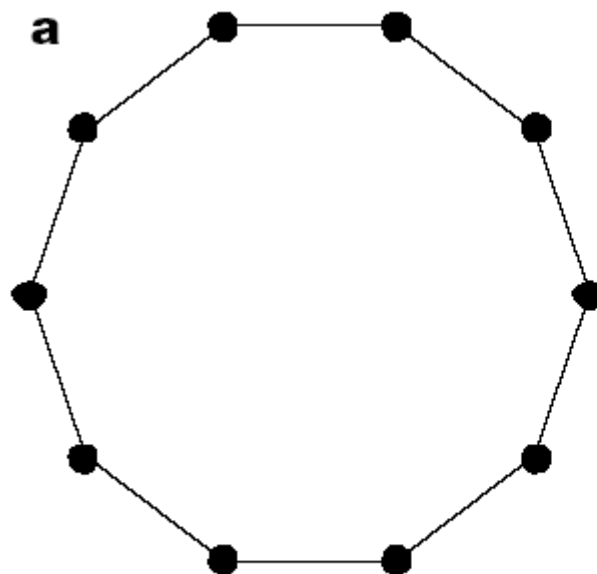
Network biology

Igor Titov ICG

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

Следующий этап постгеномной эры - исследование взаимодействий

- между генами
- между их продуктами



Сети регулярной архитектуры

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

Два параметра сети:

N – число узлов

P – вероятность связи

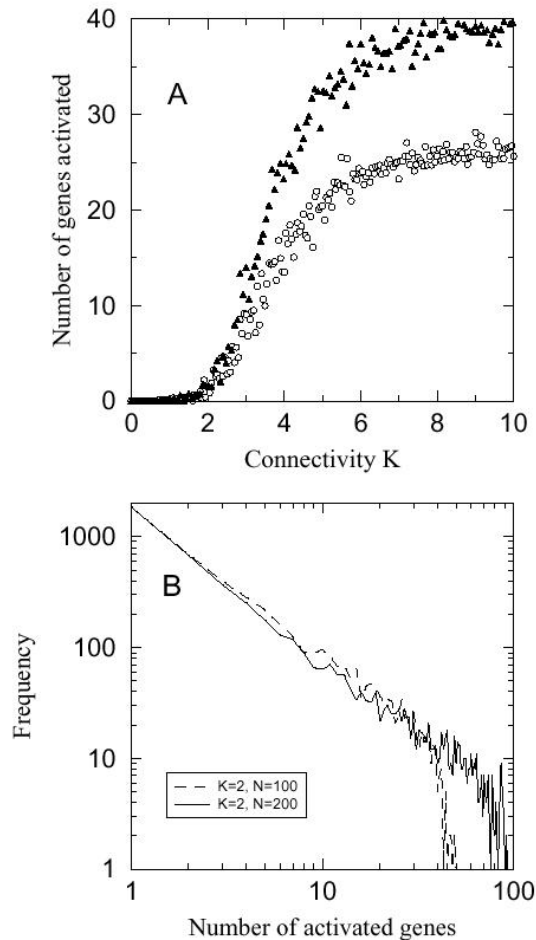
Число связей $pN(N-1)/2$

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

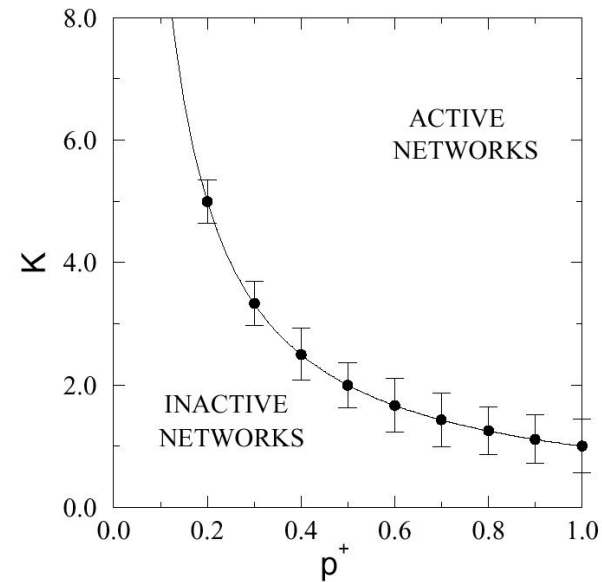
Порог перколяции $\langle k_c \rangle = 1$

Диаметр?

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

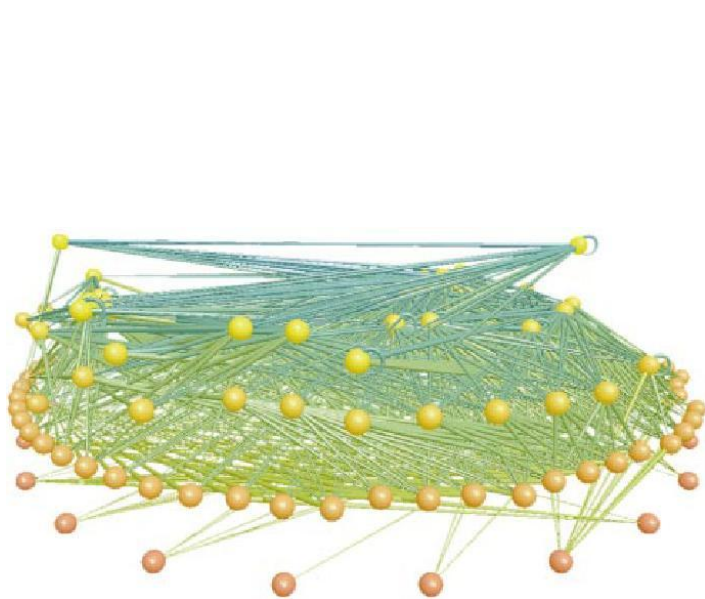


Фазовый переход в
случайной генной сети

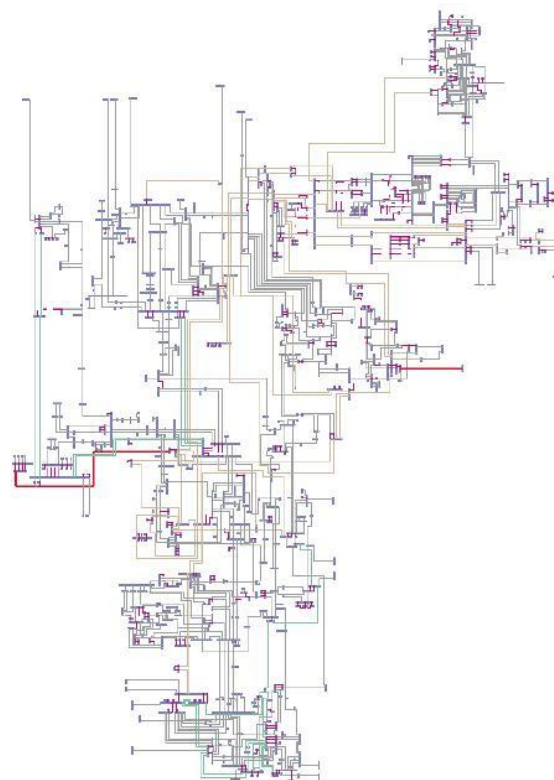


Фазовая диаграмма
случайной генной сети

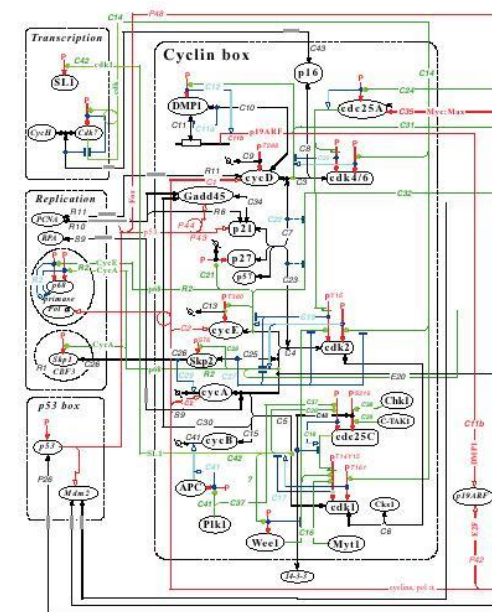
p - вероятность активации



Пищевая карта озера
Литтл Рок



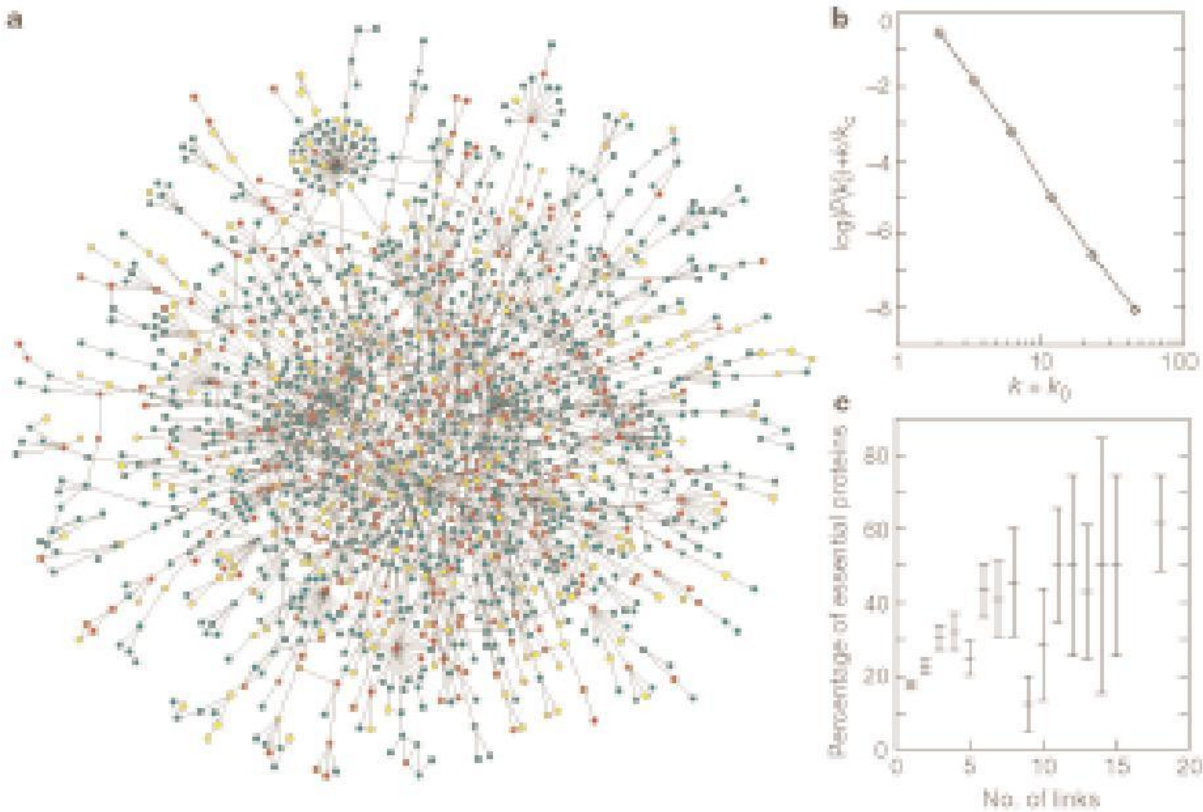
Электросеть штата Нью Йорк



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

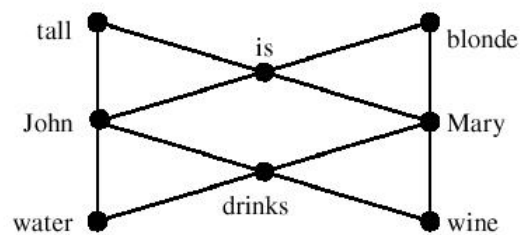
Характеризация протеомы E.Coli

- 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)

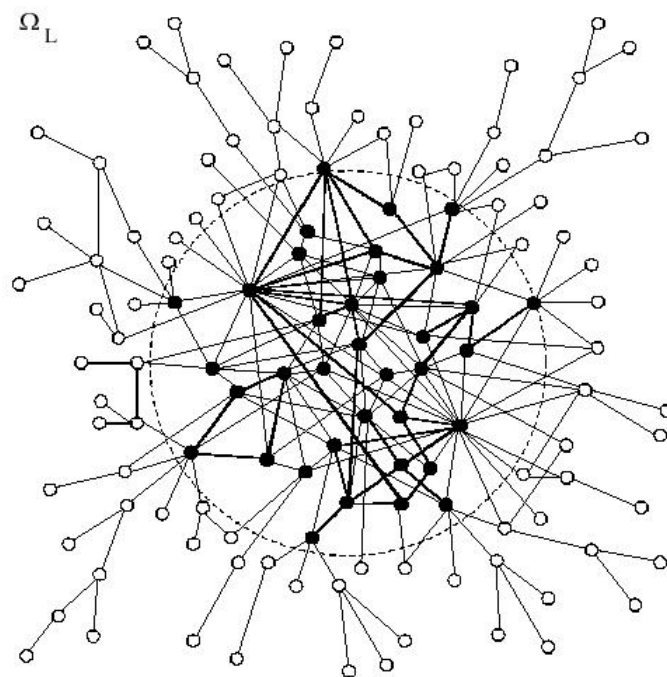
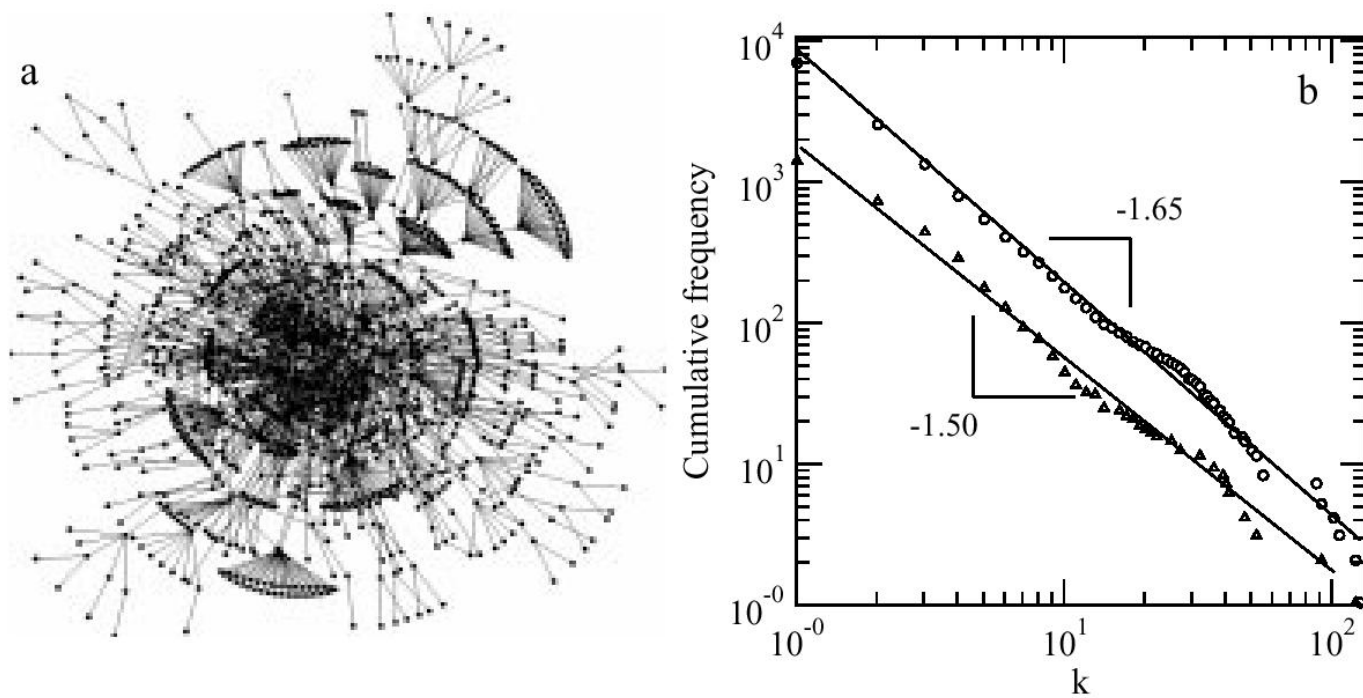


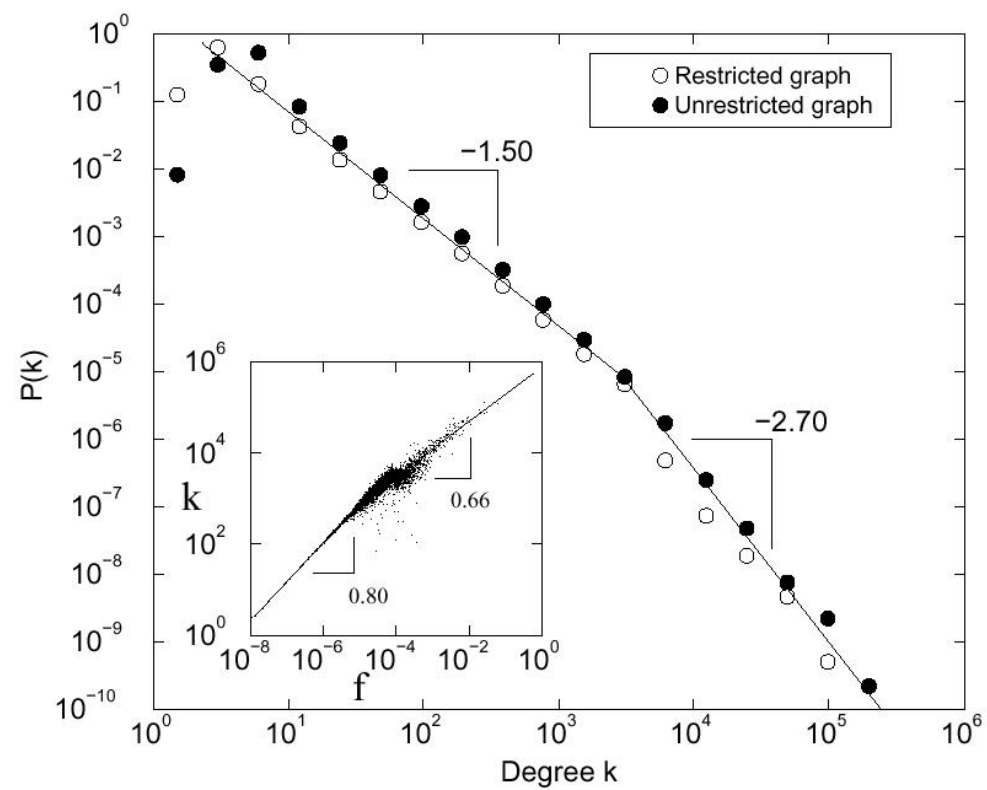
Диаграмма классов Java Development Framework 1.2



Распределение связности в словарных сетях

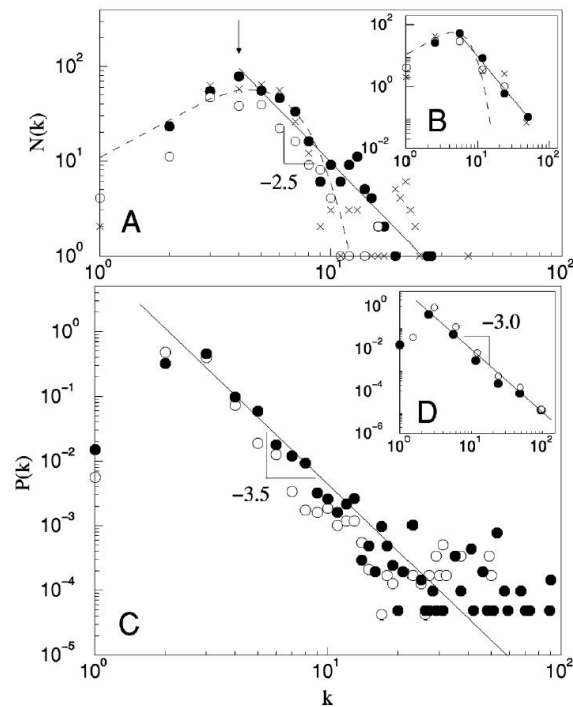
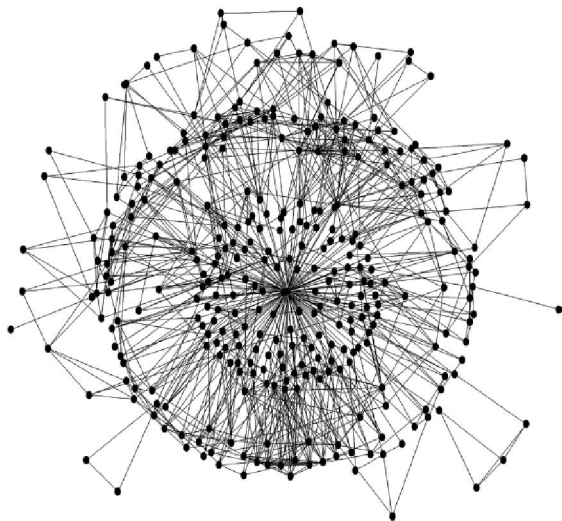
малый мир (2-3)

безмасштабное рапределение связности



Электрические устройства

Граф аналогового телевизора



Распределение связности
в малых устройствах (3
телевизора)

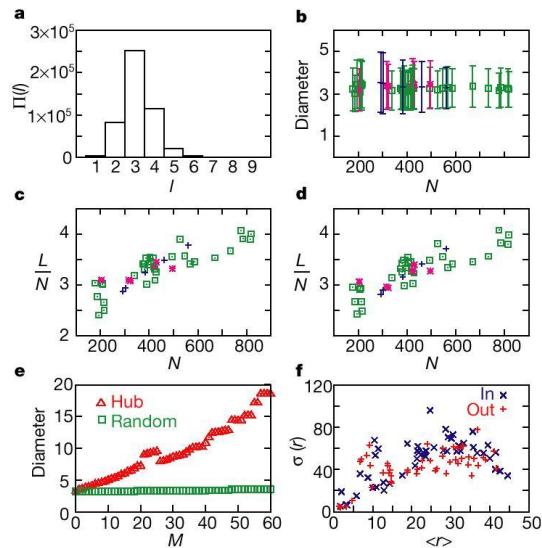
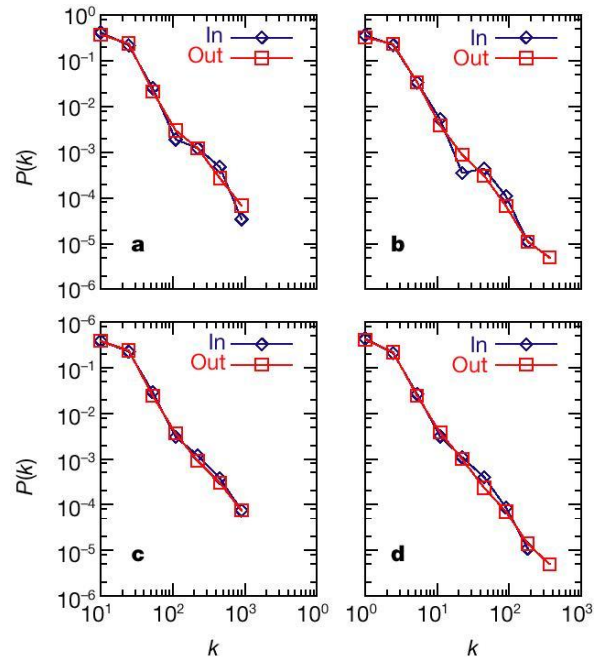
максимум при
связности=4, верхний
порог - 25-30

пунктир - пуассон

Распределение связности
в больших устройствах

верхний порог - 100

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

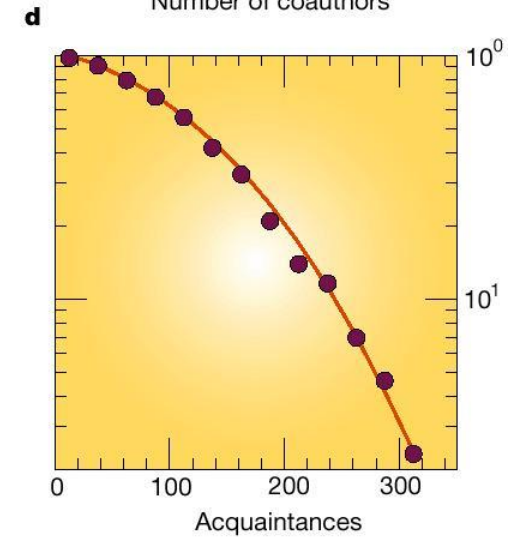
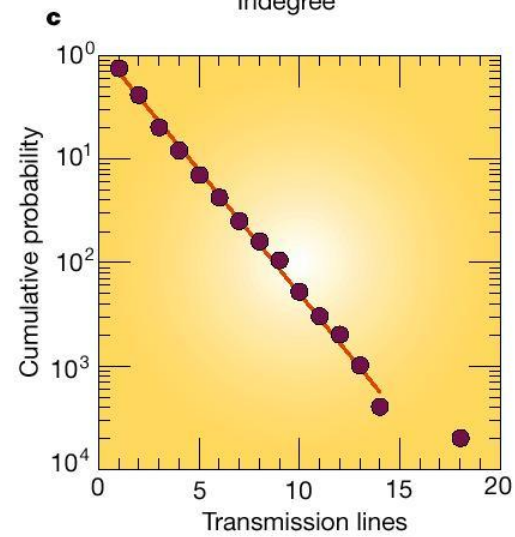
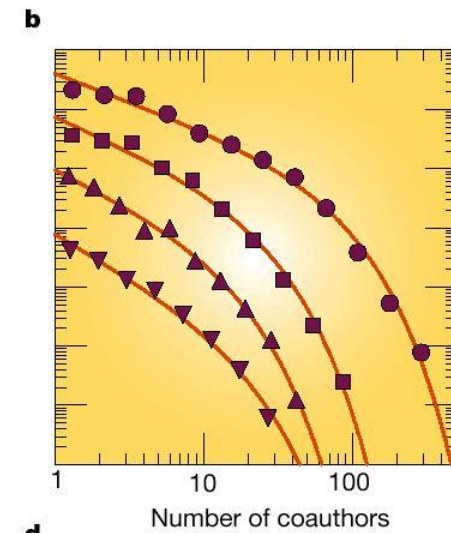
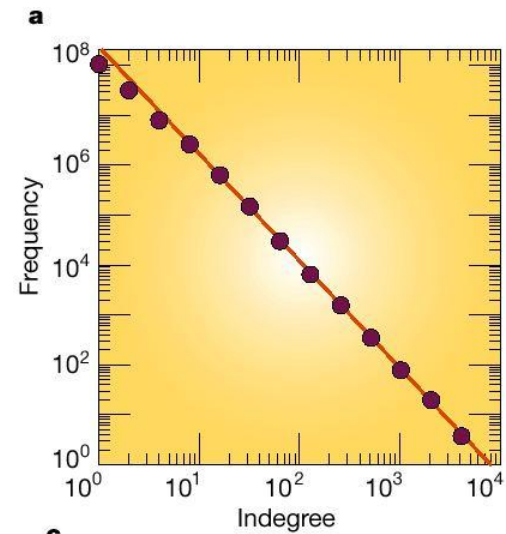
Распределение связности реальных сетей

интернет

соавторство

электросеть США и Канады

социальная сеть мормонов

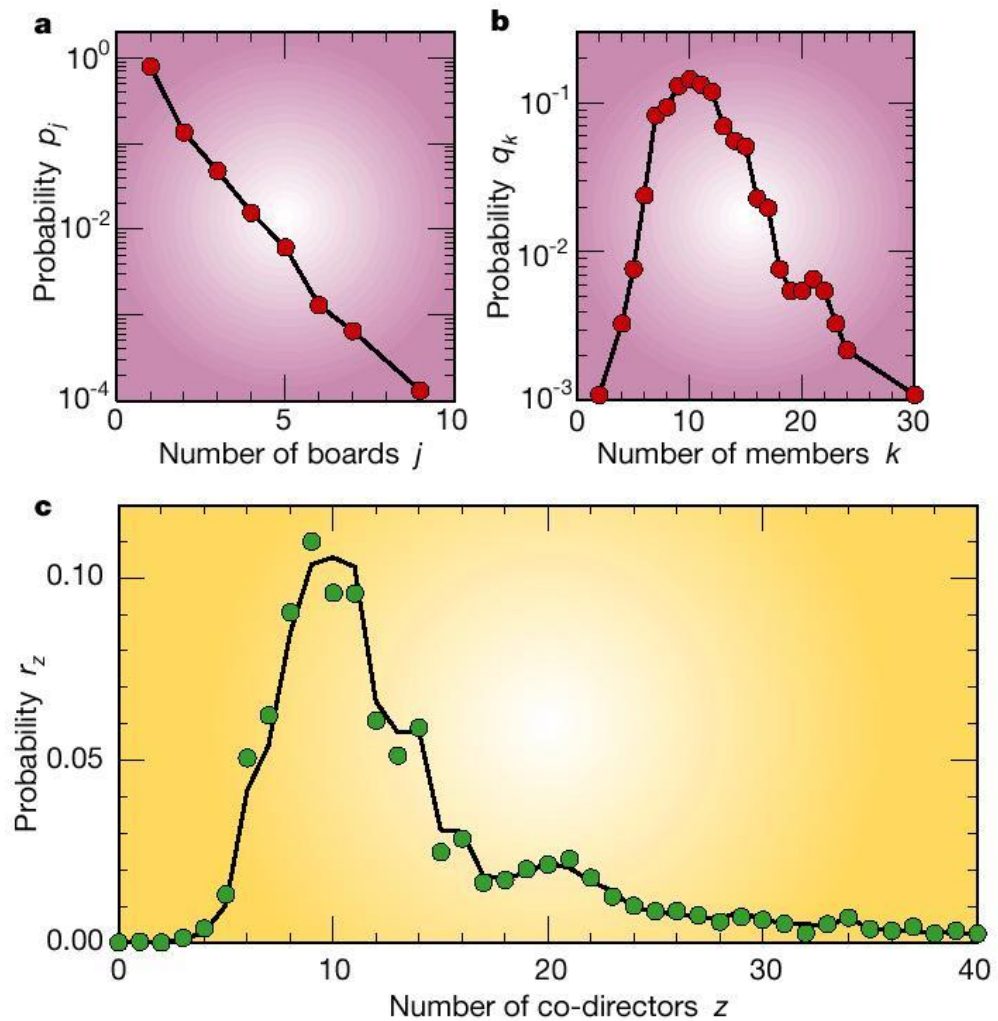


Сеть директоров корпораций

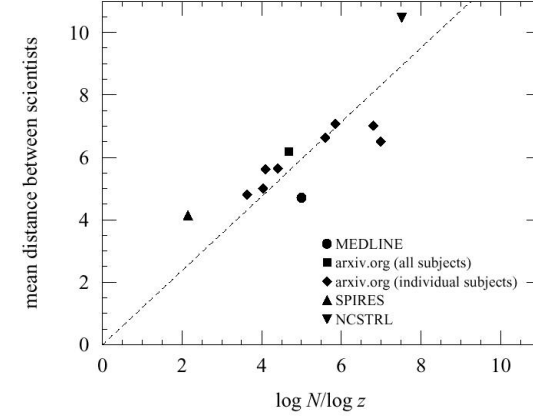
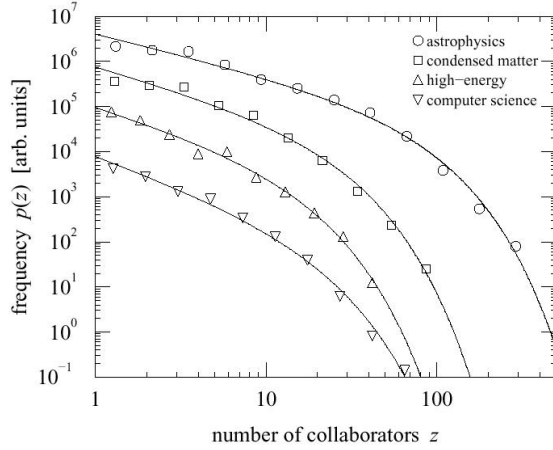
распределение числа постов на одного
директора

распределение корпораций по числу
директоров

распределение по числу со-директоров



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.

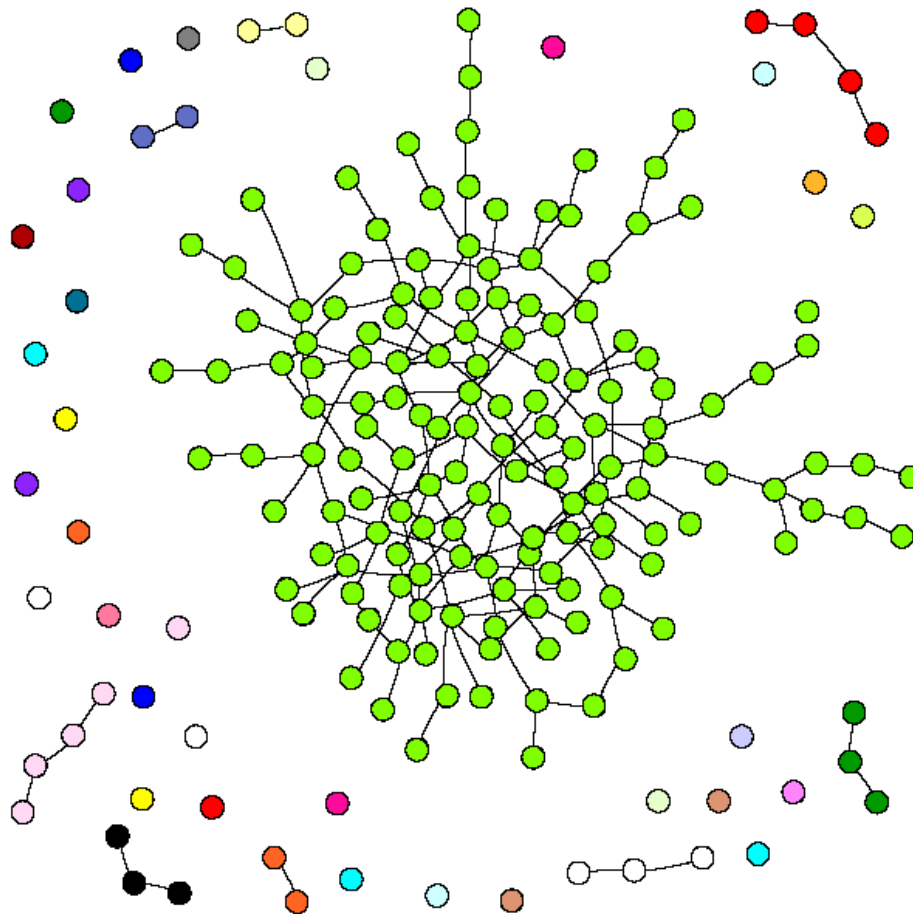
Основные структурные характеристики узлов и сетей

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 наибольшем кластере

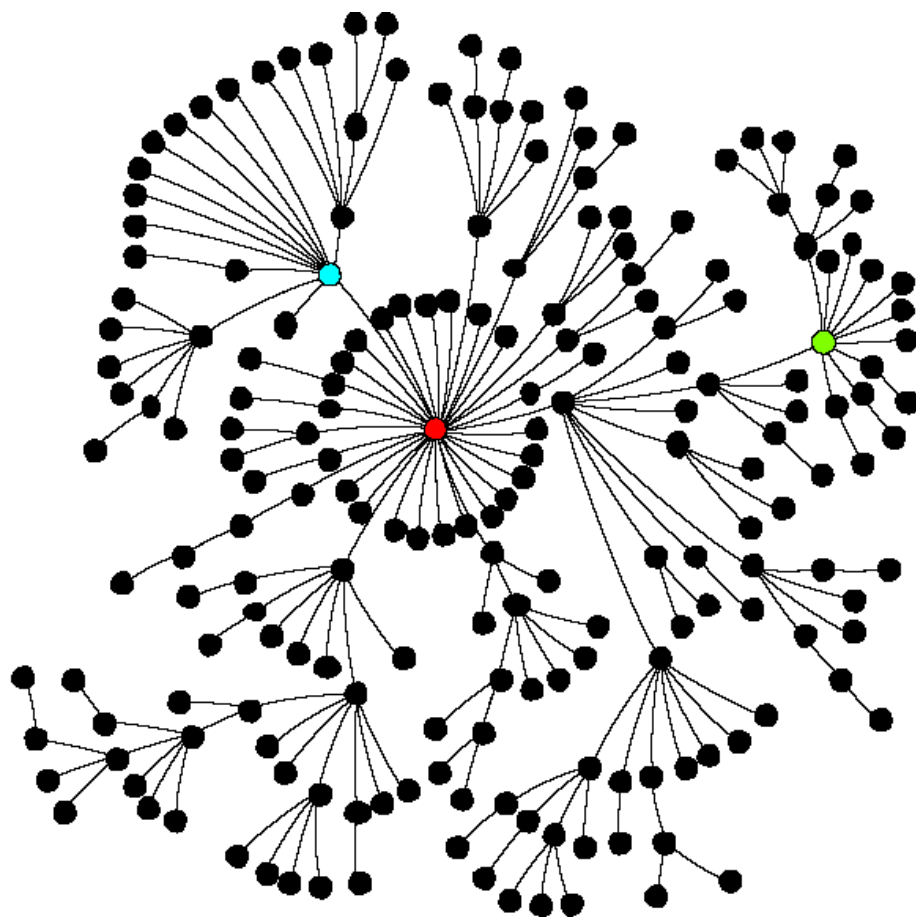


Случайный граф

Число вершин
 $N=200$

Макс связность
одной вершины
 $z=193$

Гигантская
компонента при
 $z > N/2$

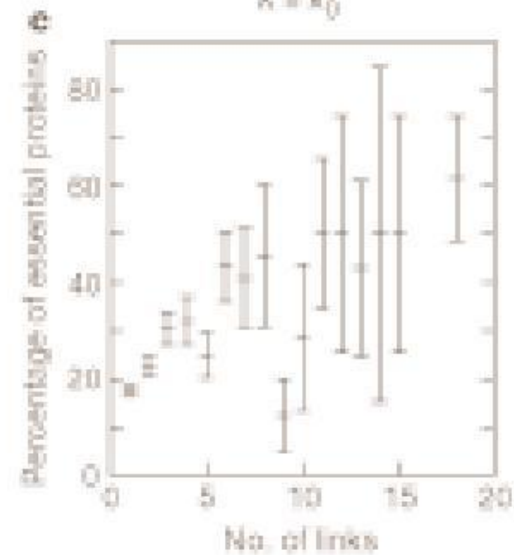
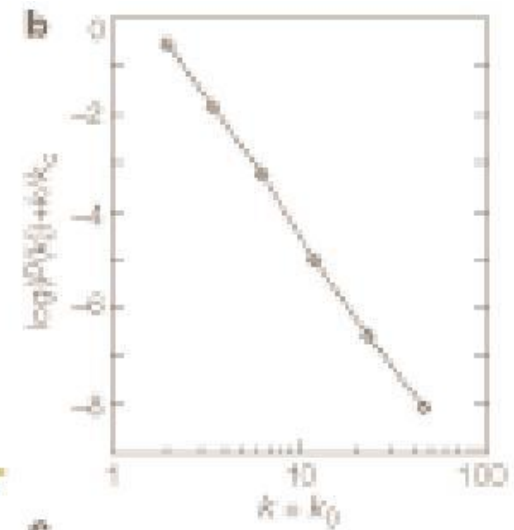
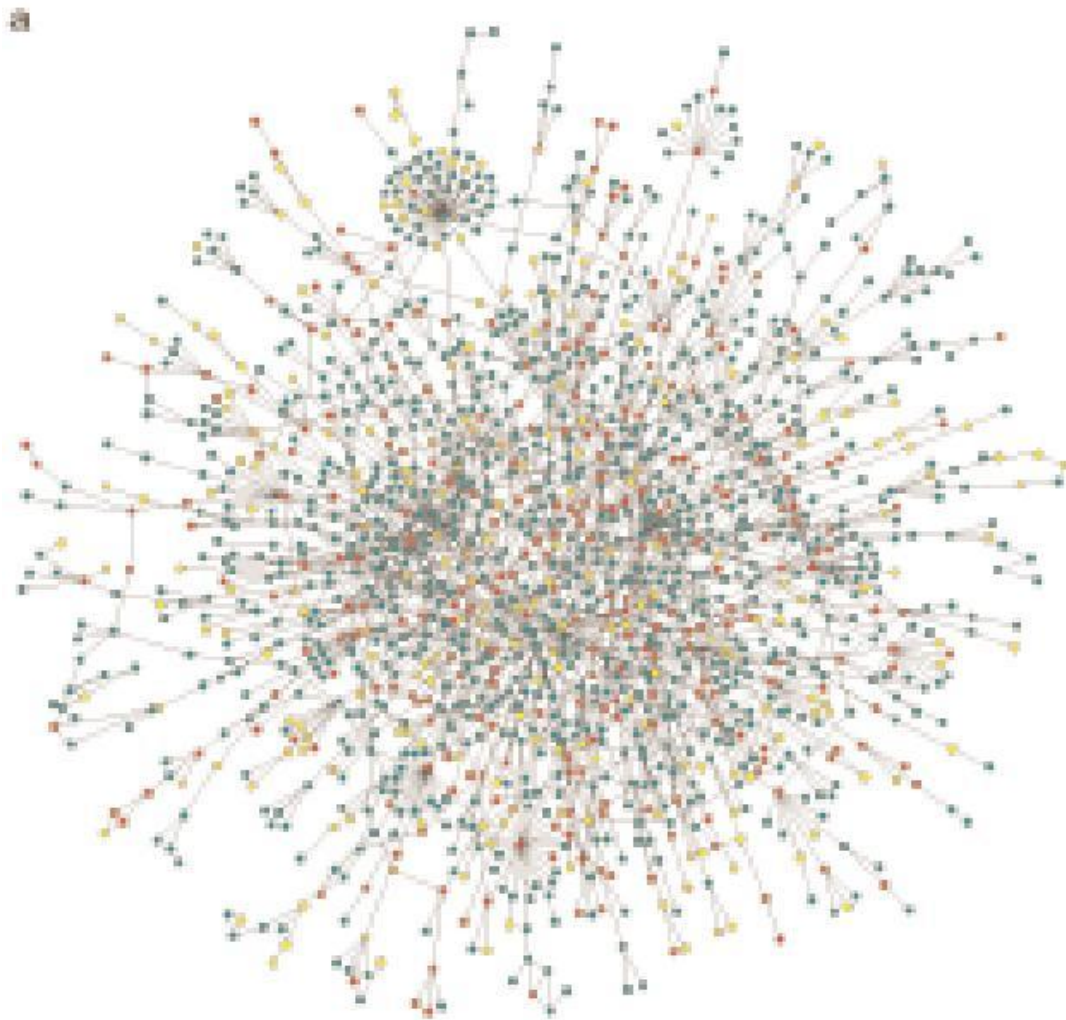


Безмасштабный
граф, полученный
присоединением с
вероятностью,
пропорциональной
текущему числу
связей

Цвет – самые
популярные

200, 100

Протеома E.Coli: 1870 белков и 2240 идентифицированных взаимодействий

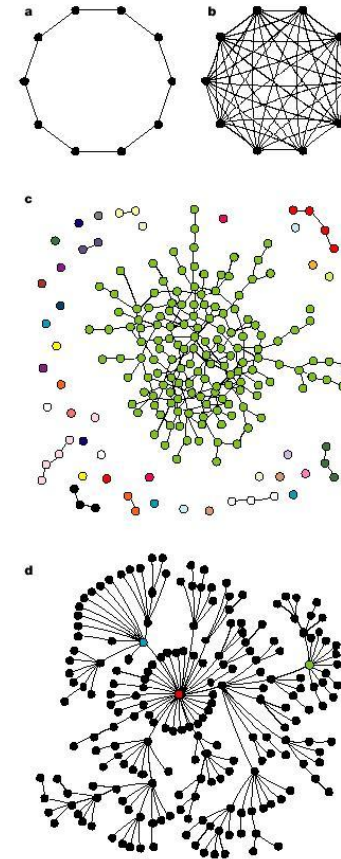
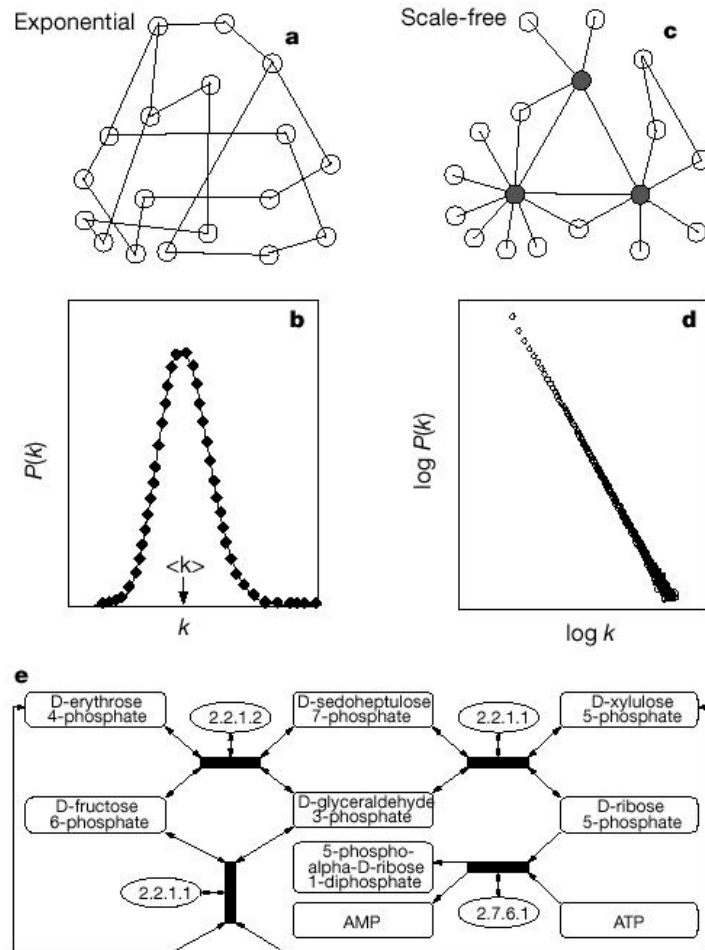


Исследованные безмасштабные сети

- 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

Структурная организация – следствие характера роста
каннибализм

Что такое безмасштабная сеть



Архитектура регулярных
и случайных сетей

Свойства безразмерных сетей

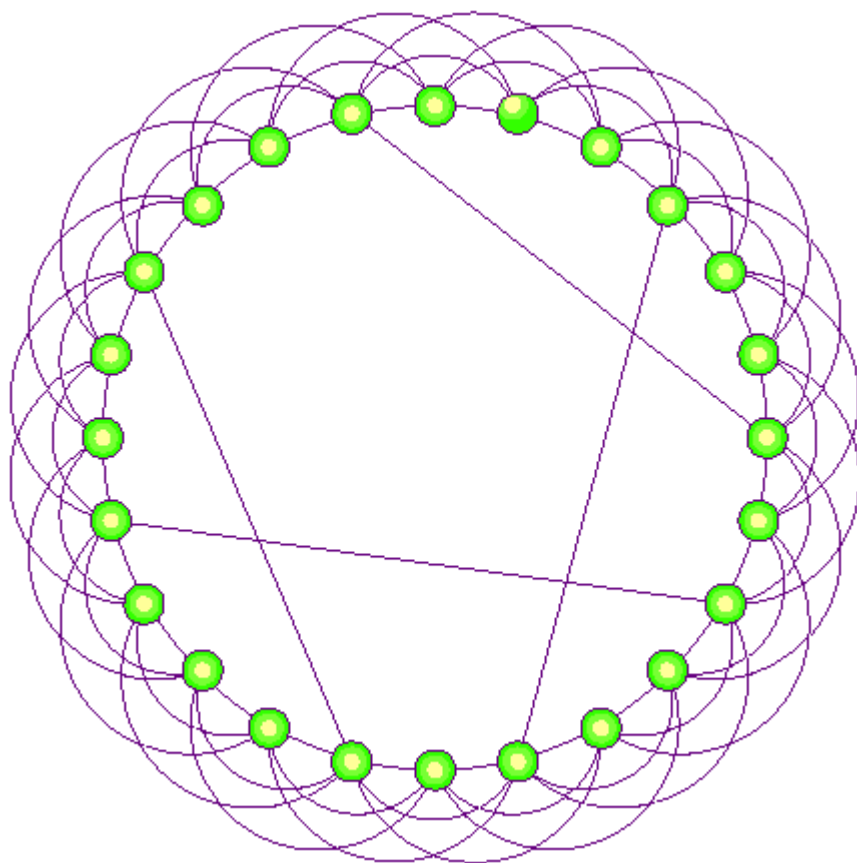
- «малый мир» (короткие пути, свойство других случайных графов)
- кластеризация (много общих соседей)
- ускорение распространения сигнала
- синхронность
- параллелизм «вычислений»

структура



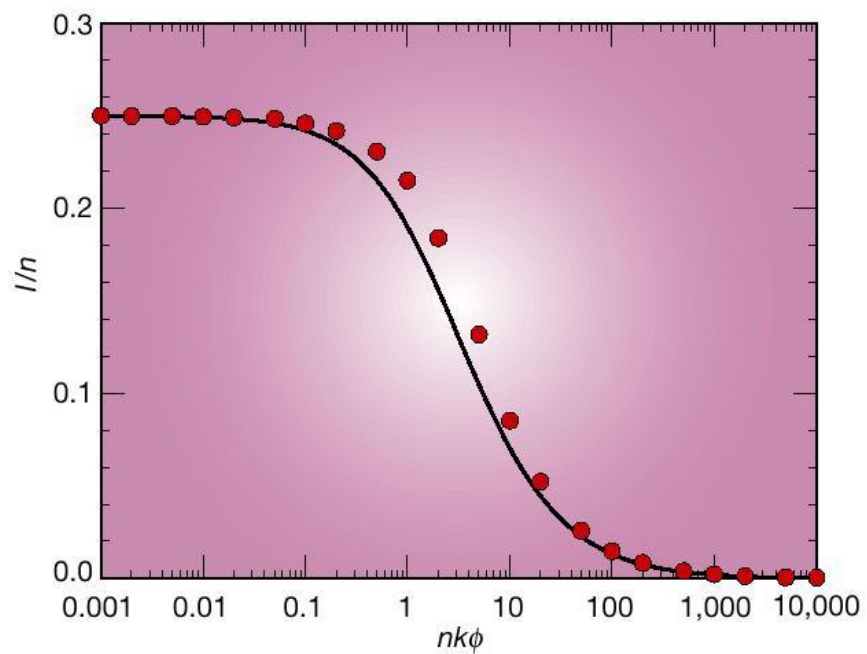
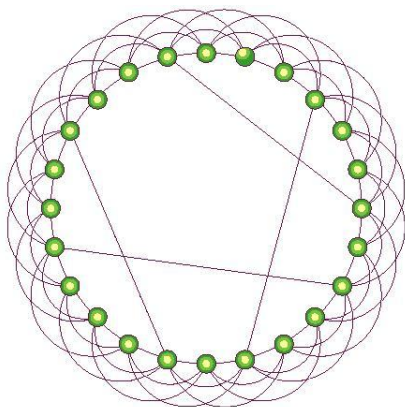
динамика

Легкость коммуникаций всей сети через каналы
+
глобальная координация потоков информации



Модель
«малого
мира»

Аналитически решаемая модель «малого мира»



Удельная средняя длина пути в
зависимости от среднего числа шунтов

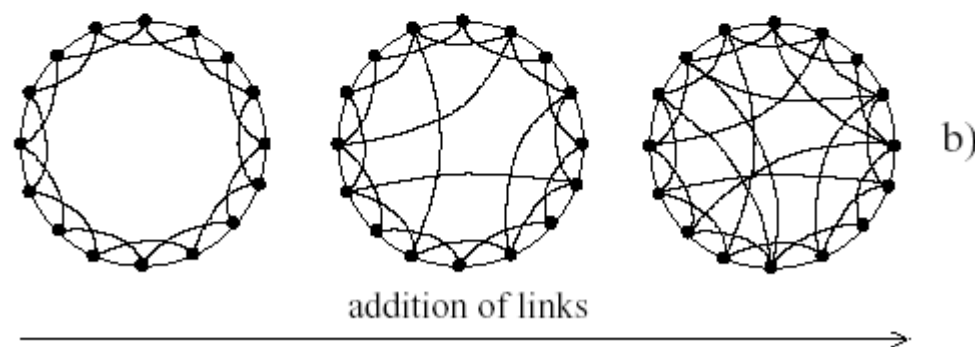
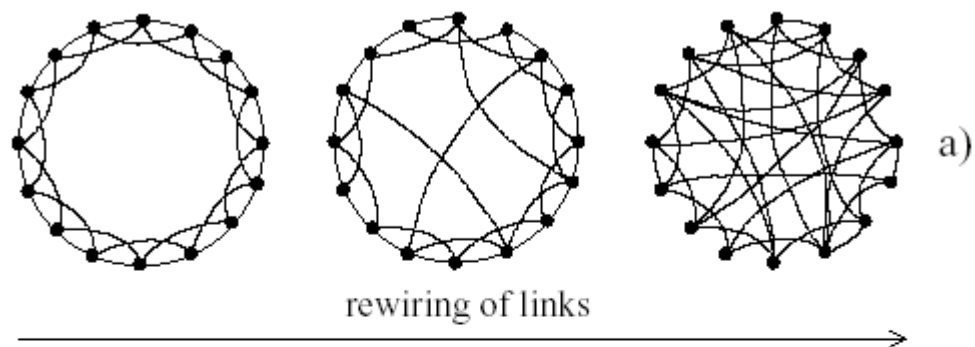


FIG. 10. Small-world networks in which the crossover from a regular lattice to a random network is realized. (a)

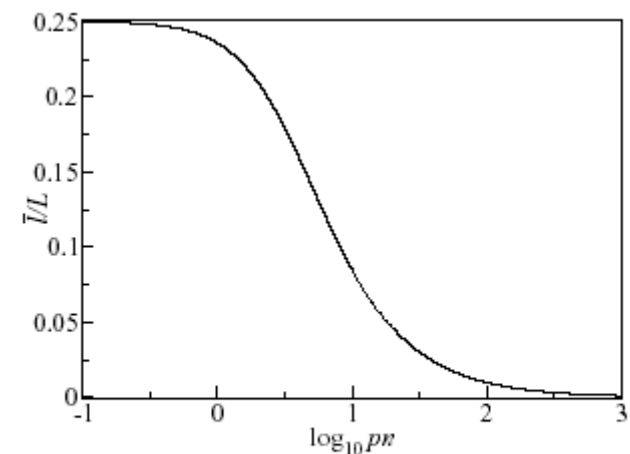


FIG. 17. The normalized average shortest-path length $\bar{l}/L$ of the “smallest-world” network vs. the number $\rho = pL$ of added edges.

Предпочтение в связывании

Popularity is attractive!

Следствие: The oldest is the richest

Безмасштабные сети – идея предпочтительного линкинга
(популярность привлекательна)

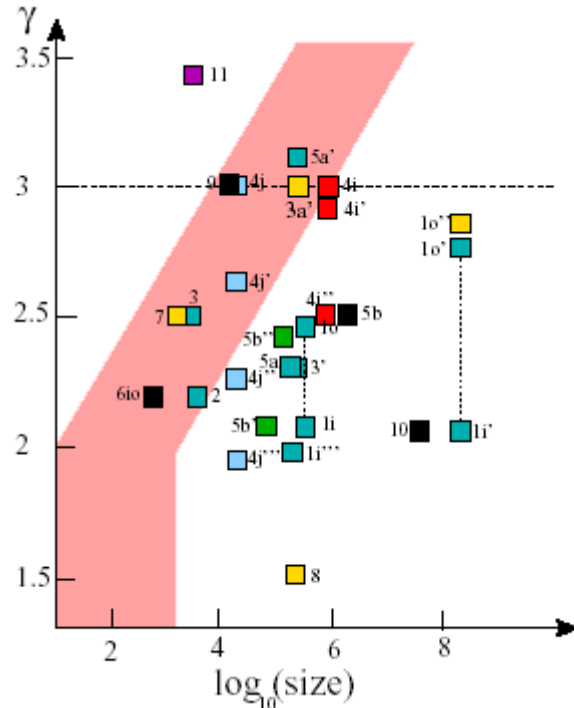
Механизмы: новые вершины

-добавление связей между старыми (чем старше, тем богаче)

-ГА

-старение

Показатели безмасштабных сетей



Ускоренный рост – число ребер растет быстрее
числа вершин (Интернет, www, цитирование,
соавторство)

Умирающие сети

Атаки и
повреждения

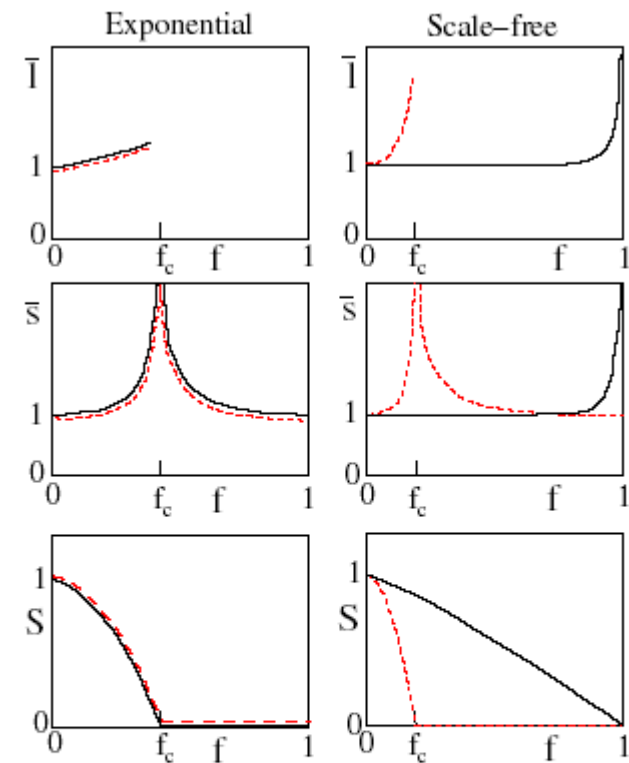


FIG. 39. Schematic plots of the effect of intentional and random damage (attack and failures) on the characteristics of exponential undirected networks and scale-free undirected networks with exponent $\gamma \leq 3$ [59]. The average shortest path between vertices, $\bar{l}$, the size of the largest connected component, $\bar{s}$, and the average size of isolated clusters, S , are plotted vs. the fraction of removed vertices $f \equiv 1 - p$. The networks are large. The solid lines show the effect of the random damage. The effect of the intentional damage is shown by the dashed lines. For the exponential networks, both kinds of damage

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?

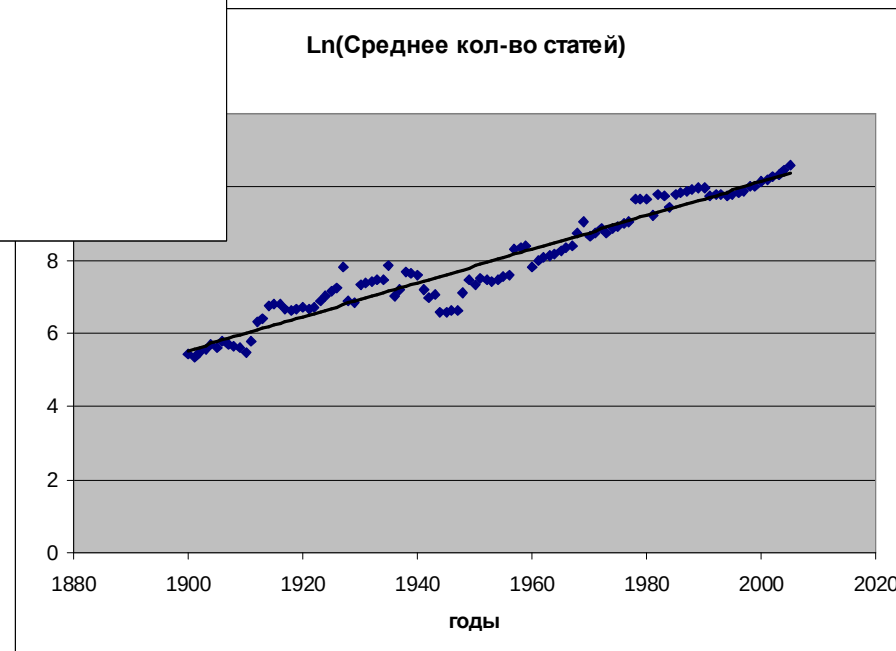
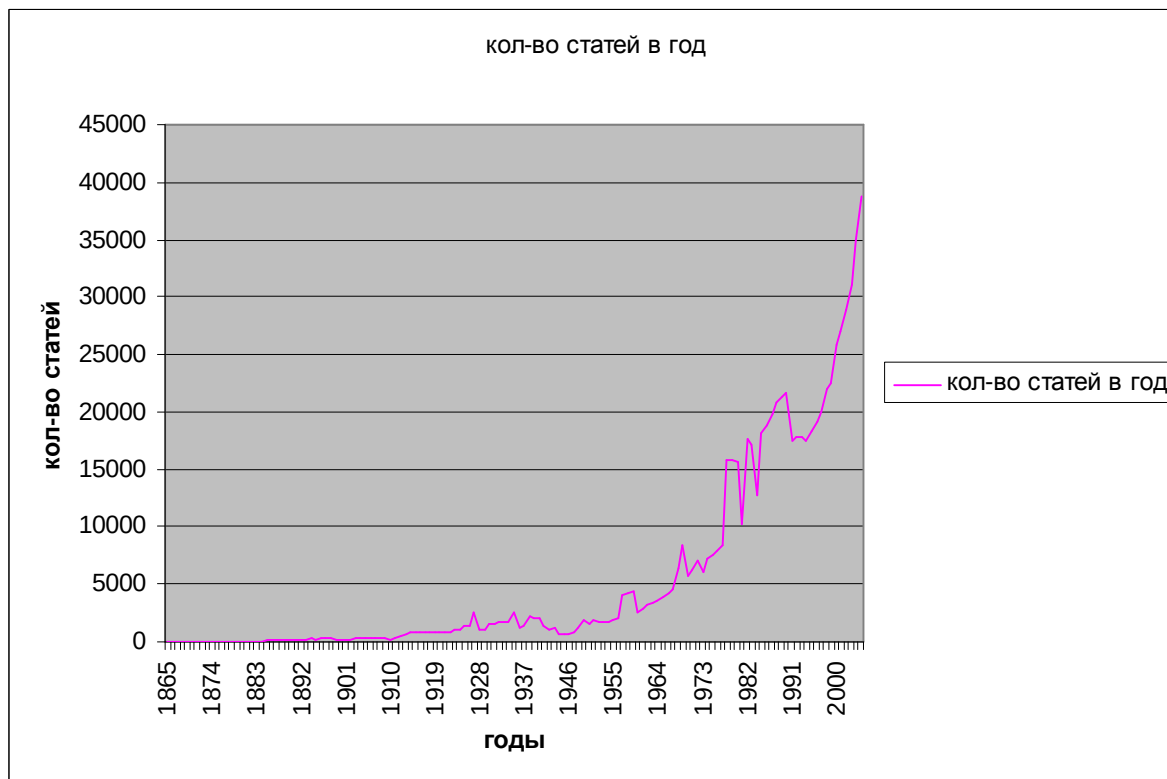
Эпидемии:

Перколяционный характер для *экспоненциальных сетей*

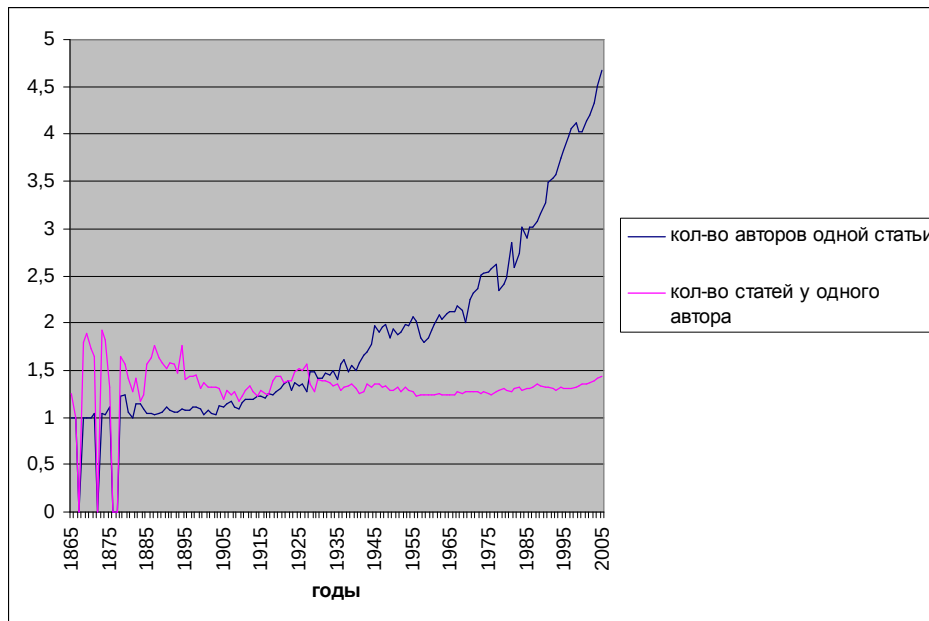
Нулевой порог для *безмасштабных сетей* с «толстыми хвостами», пандемии

Иммунизация ядра

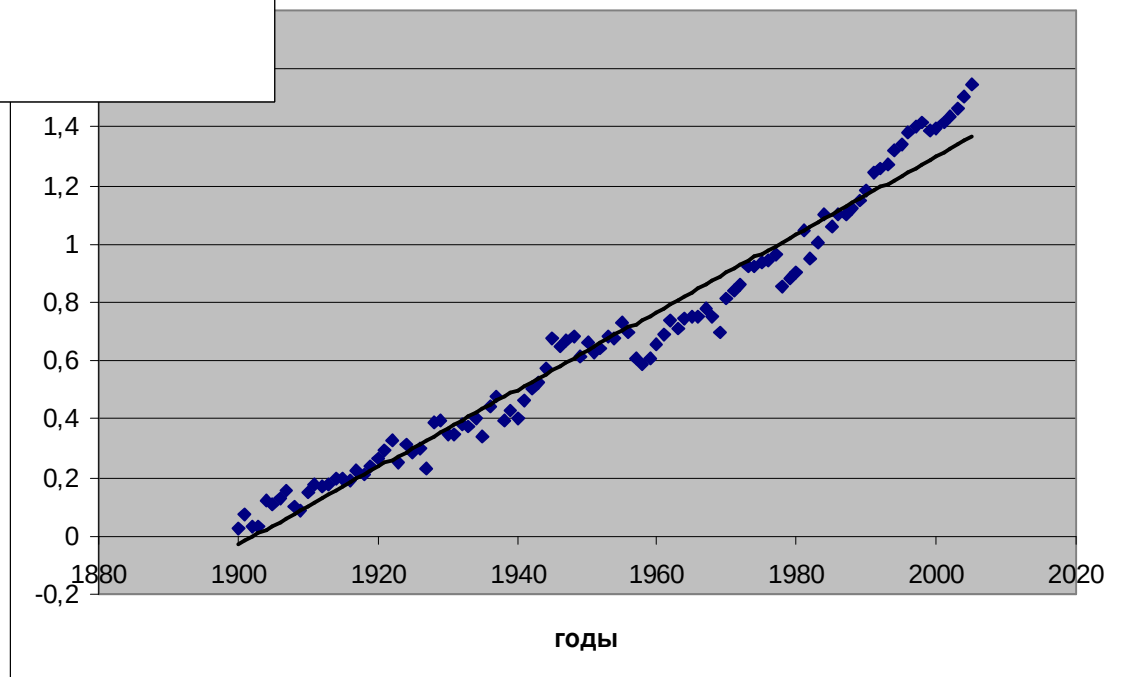
Динамика роста количества научных публикаций в PubMed



Динамика роста числа соавторов

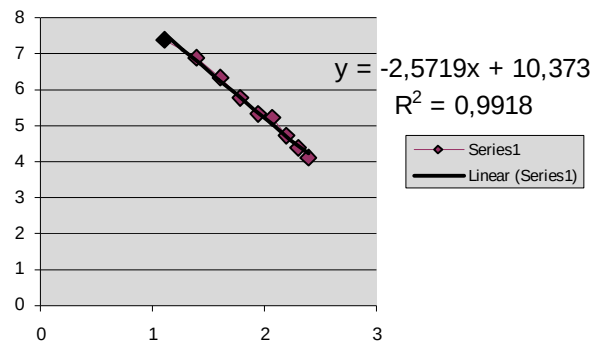


Ln(Среднее число соавторов 1 статьи)

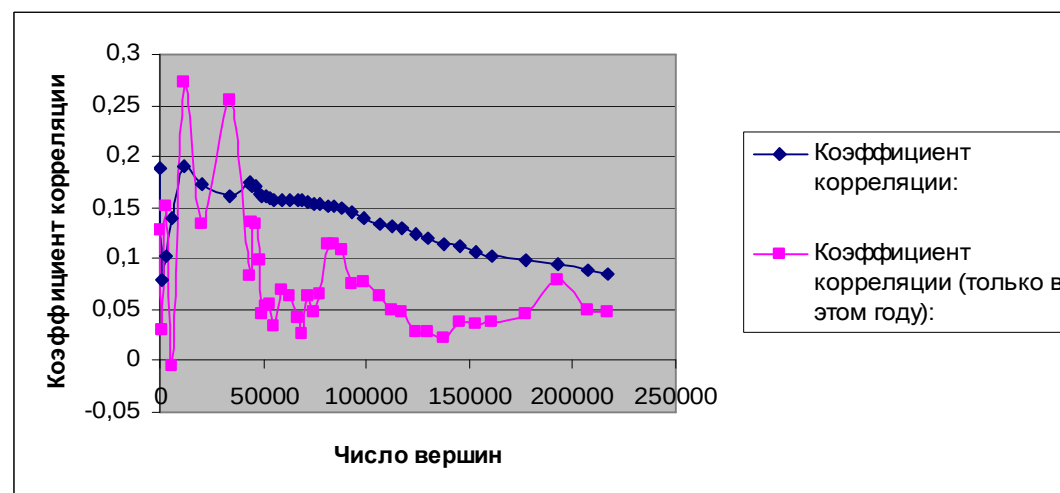
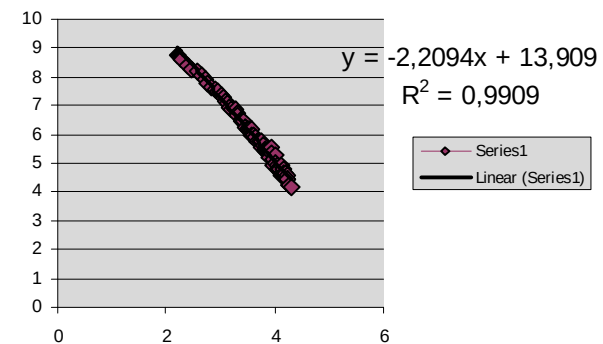


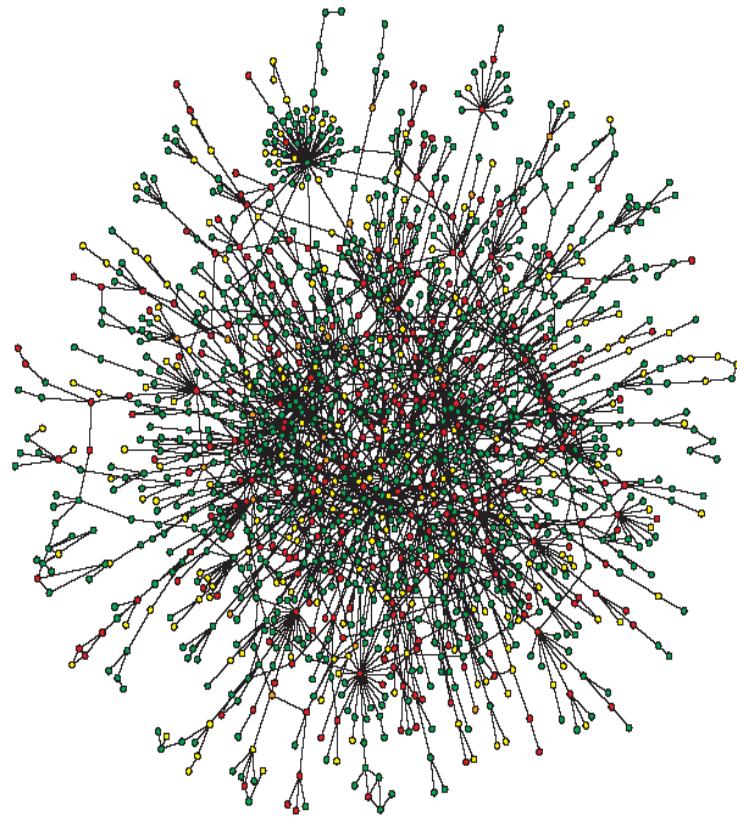
Инвариантность безмасштабной архитектуры сети соавторства

1960



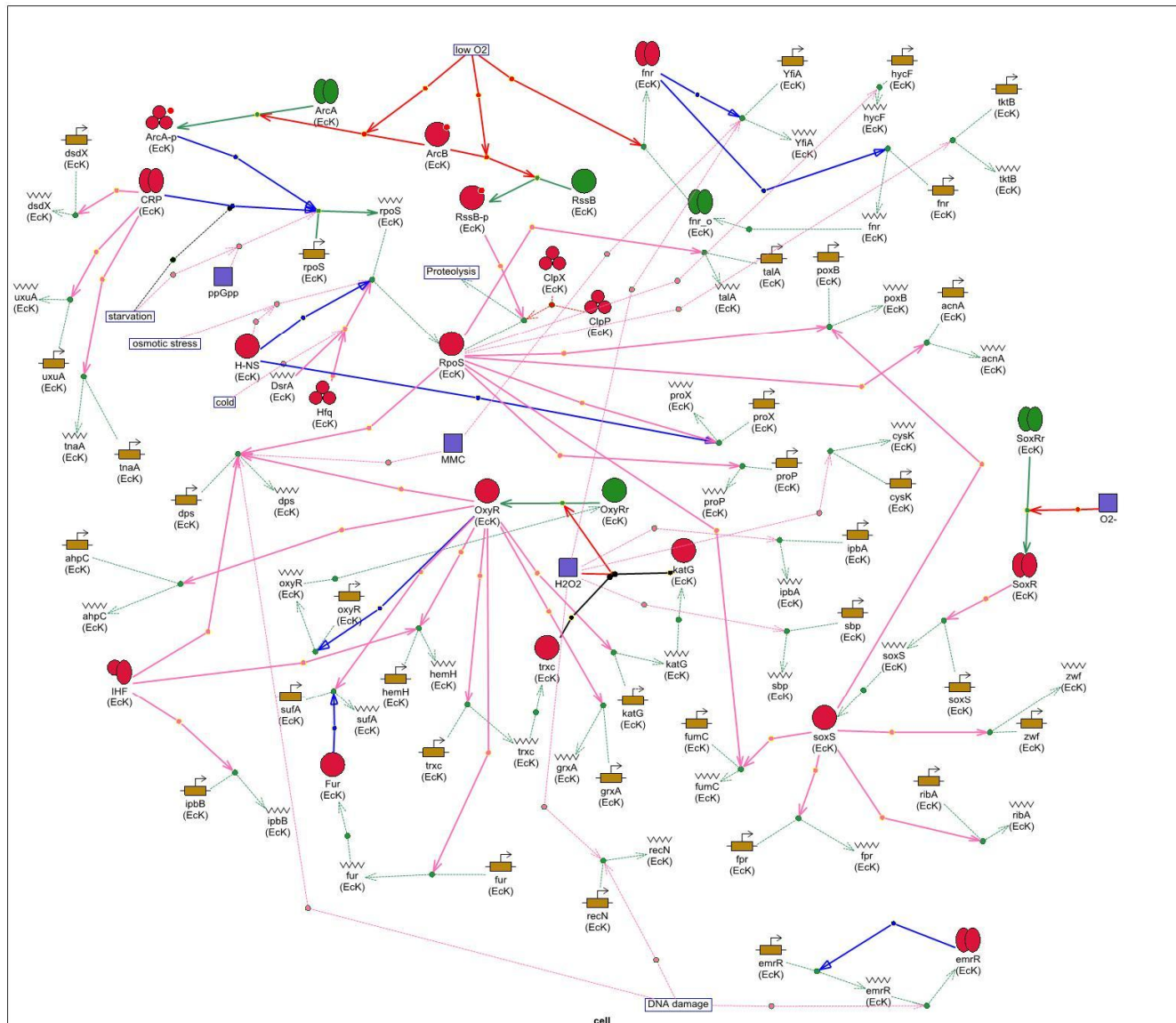
2005





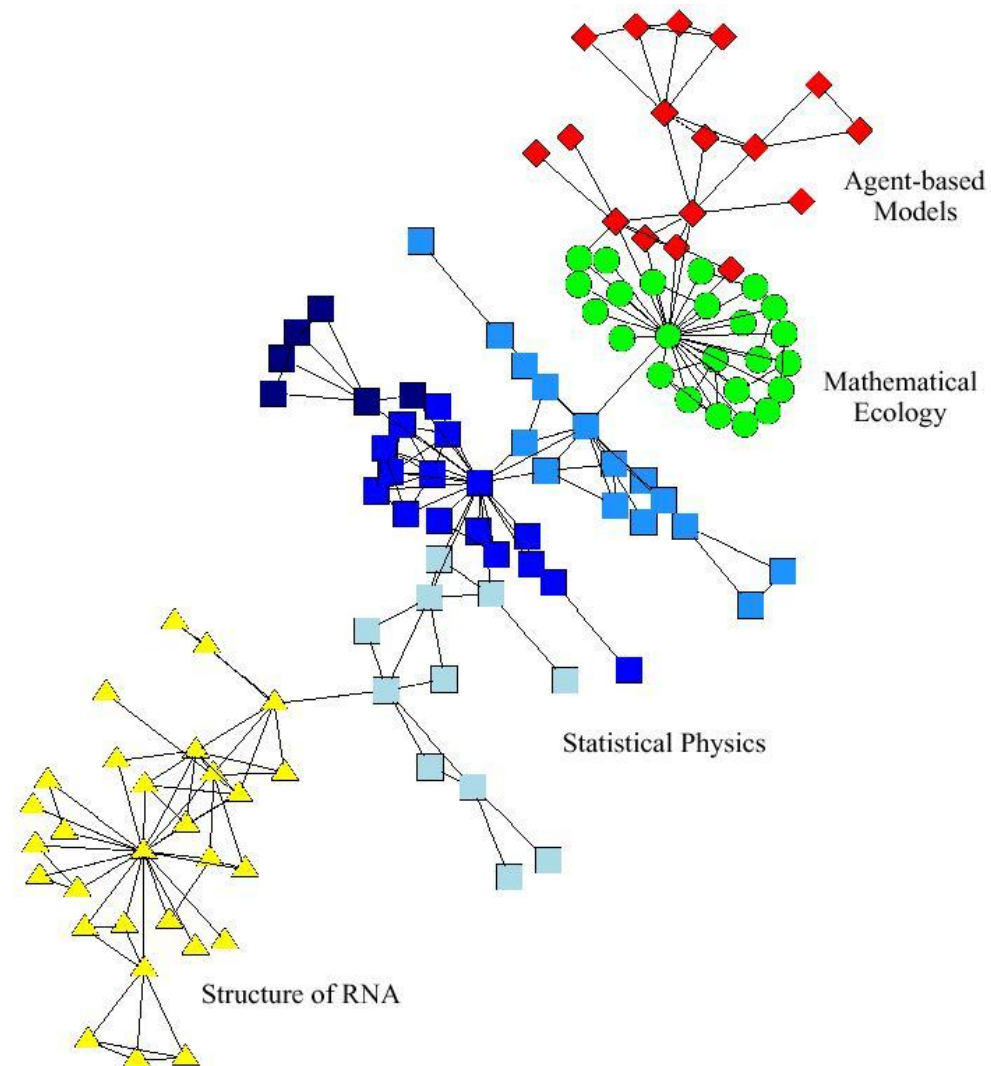
Редукция сетей

- Разбиение на кластеры
- Поиск мотивов
- Поиск ключевых звеньев



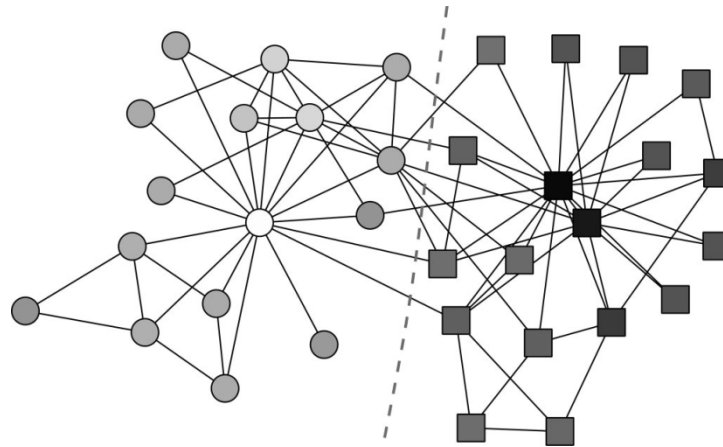
Что делать с сетями промежуточного размера?
геновая сеть реакции E.coli на стресс

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



Алгоритм “жадной” оптимизации модульности

- Сеть разбивается на два сообщества, после чего эти два сообщества разбиваются на еще более мелкие модули, при этом модульность стремится к единице, используя «жадную» оптимизацию
- Метод представляется как самый быстрый для разбиения графа на модули очень больших сетей.



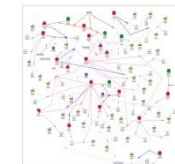
Алгоритм “Walktrap” на основе случайных блужданий

- Случайные блуждания по графу имеют тенденцию «попадать в ловушку» в тесно связанные части, называемые сообщества
- Благодаря случайным блужданиям, определяется структурное подобие между вершинами и сообществами, таким образом определяя расстояние
- Это расстояние используется в иерархическом алгоритме объединения в кластеры, что и дает конечный результат.

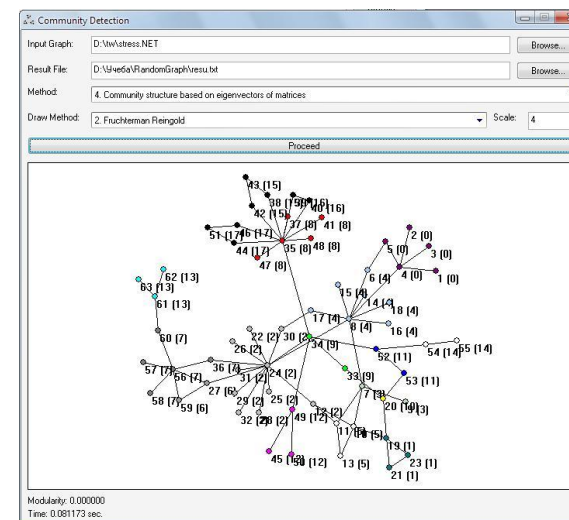
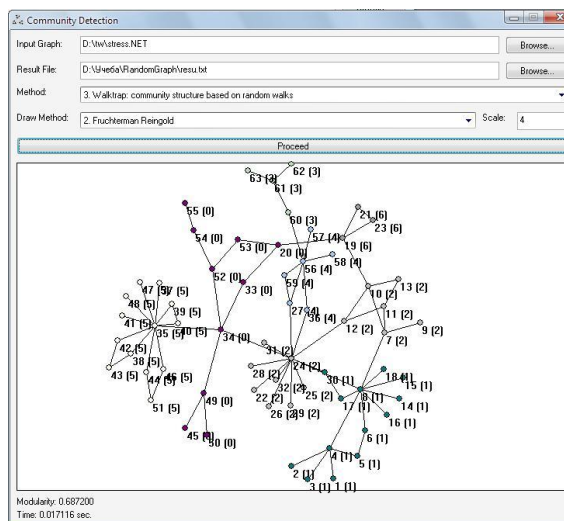
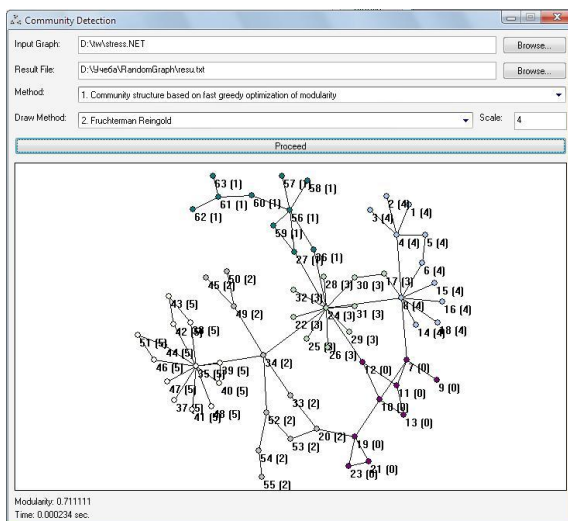
Алгоритм на основе определения смежности

- Вместо того, чтобы пробовать строить величину, которая говорит нам, какие ребра являются центральными к сообществам, находятся ребра, которые являются менее центральными и находятся «между» сообществами
- Вместо того, чтобы строить сообщества, добавляя самые сильные ребра к первоначально пустому набору вершин, сообщества строятся прогрессивно удаляя ребра из оригинального графа.

Сравнение алгоритмов



- Результаты вычислений на графе генной сети реакции E.coli на стресс:

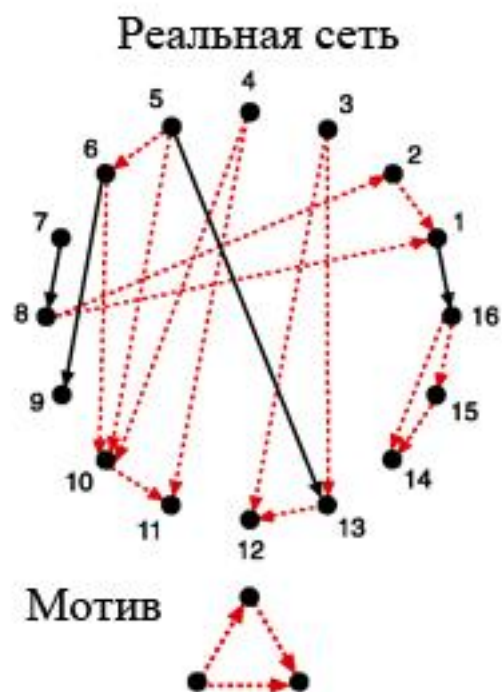


Membership: 4 4 4 4 4 0 4 0 0 0 0 4 4 3 4 0 2 0 3 0 3 3 3 1 3 3 3 3 2 2 5 1 5 5 5 5 5 5 5 2 5 5 5 2 2 5 2 2
2 2 1 1 1 1 1 1 1 1

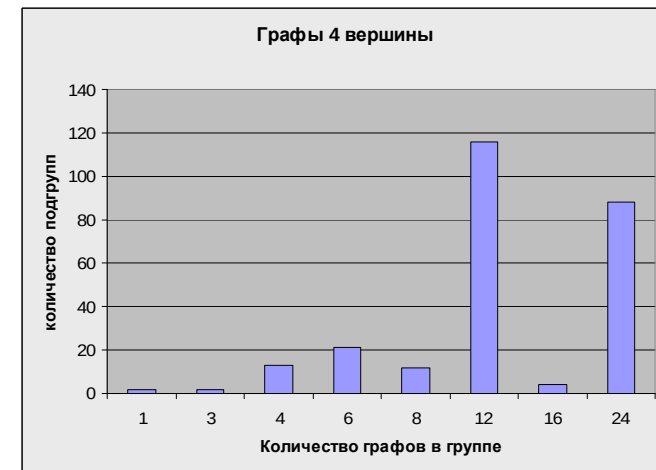
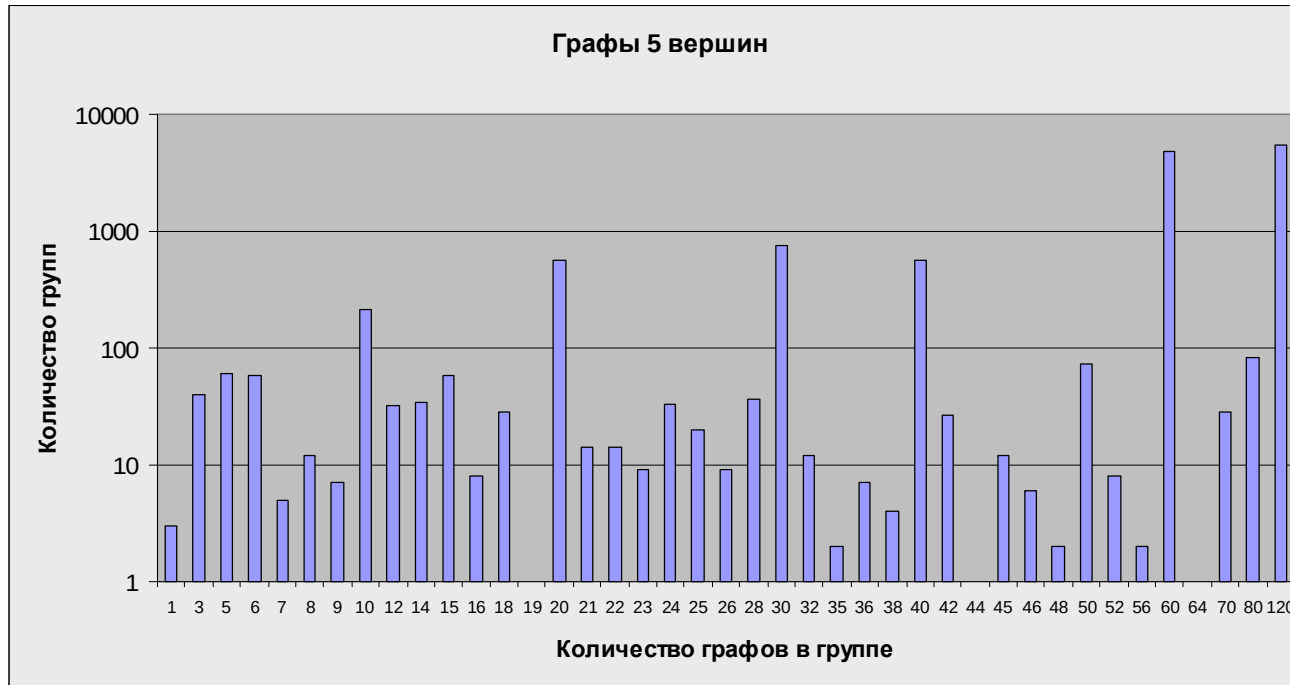
Membership: 1 1 1 1 1 1 2 1 2 2 2 2 2 1 1 1 1 1 6 0 6 2 6 2 2 2 4 2 2 1 2 2 0 0 5 4 5 5 5 5 5 5 5 0 5 5 5 0 0 5 0 0
0 0 4 4 4 4 3 3 3 3

Membership: 0 0 0 0 0 4 3 4 3 5 5 2 5 4 4 4 4 4 1 1 0 1 2 1 2 2 2 6 2 2 2 2 9 9 8 7 8 15 16 16 8 15 15 17 12 17 8

Мотив – неслучайно часто встречающийся подграф



Распределение числа графов по изоморфным группам



03 [2/7]

Original FREQ : 43.411

Random Mean FREQ : 0.02985



Random Deviation : 0.0067

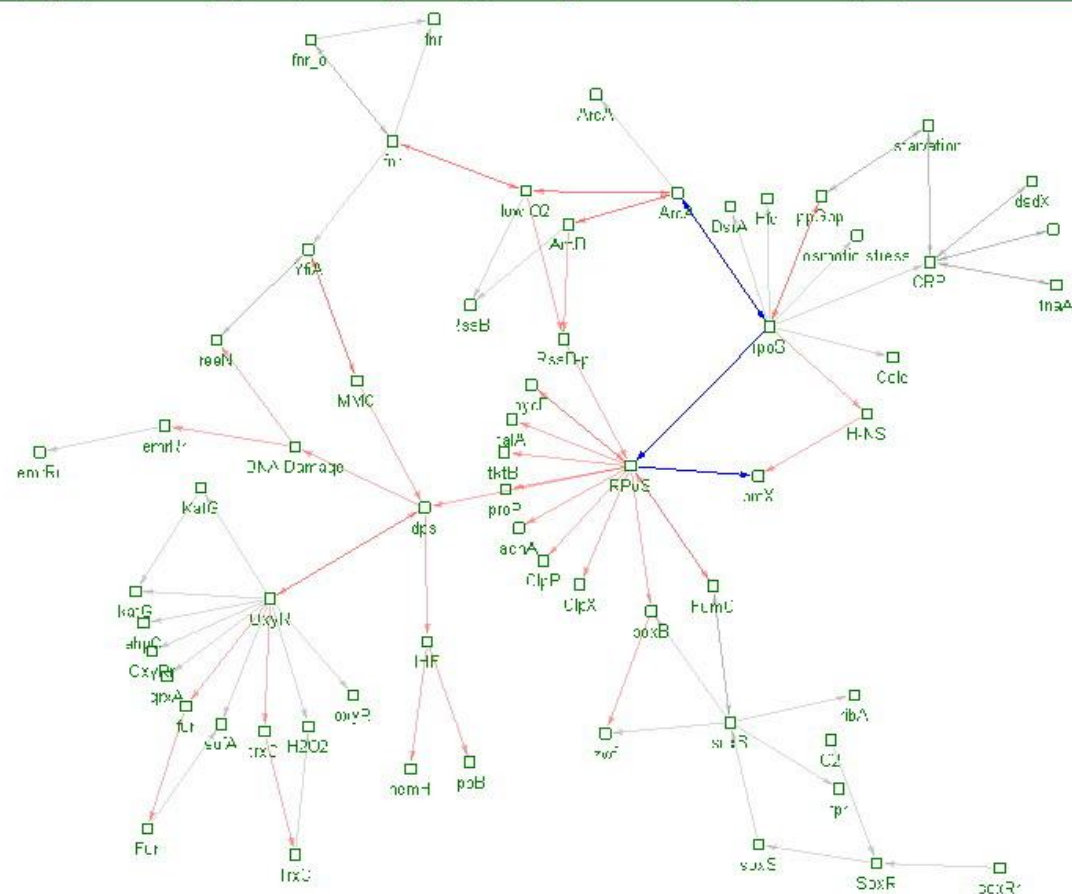
Z-Score : 64.95634

P-Value : 0.00000

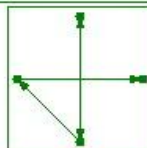
- Z-scores is calculated as

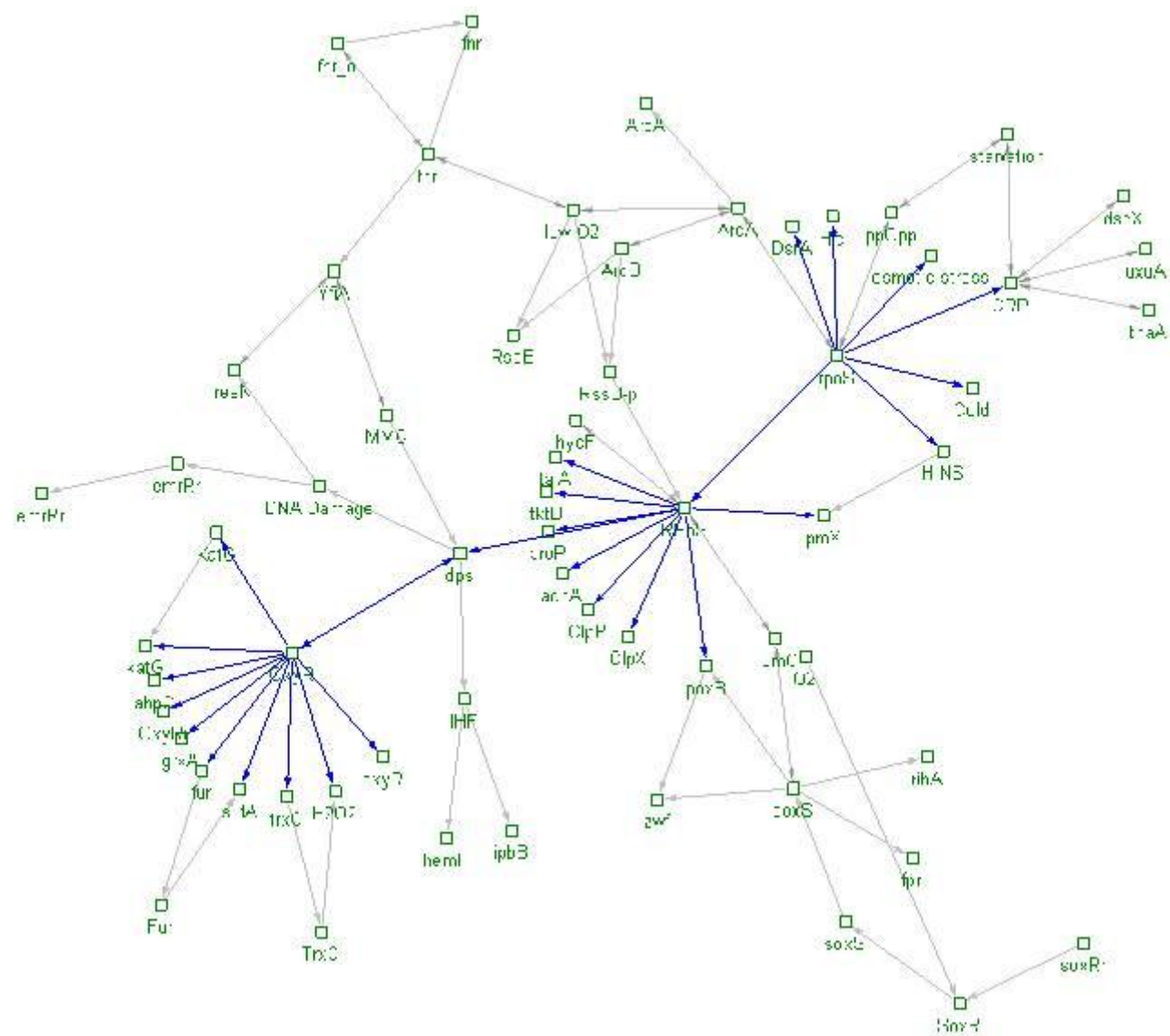
$$Z = (C_{real} - \langle C_{rand} \rangle) / Var_{rand}$$

- where C_{real} is the concentration in the real network, $\langle C_{rand} \rangle$ and Var_{rand} are the mean and SD in the randomized networks.

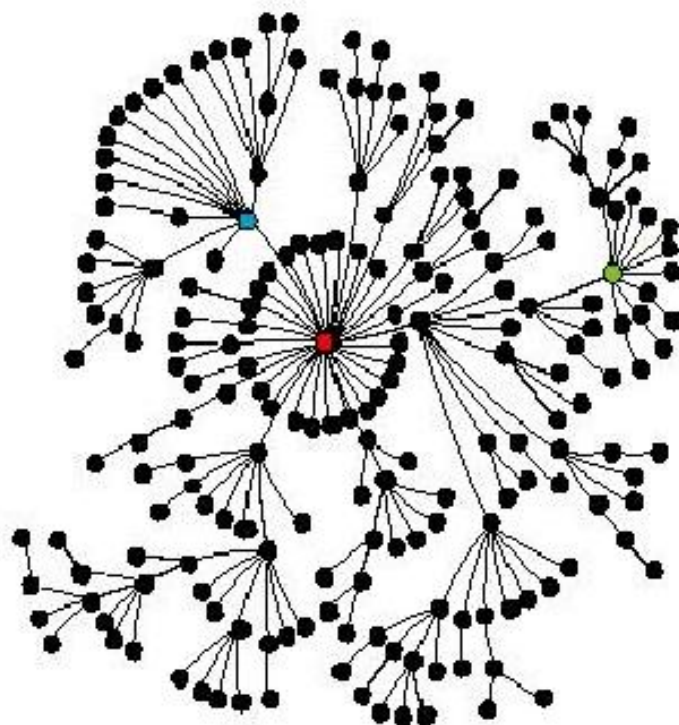


Random Mean FREQ : 0.00014

**Motif 18/36**

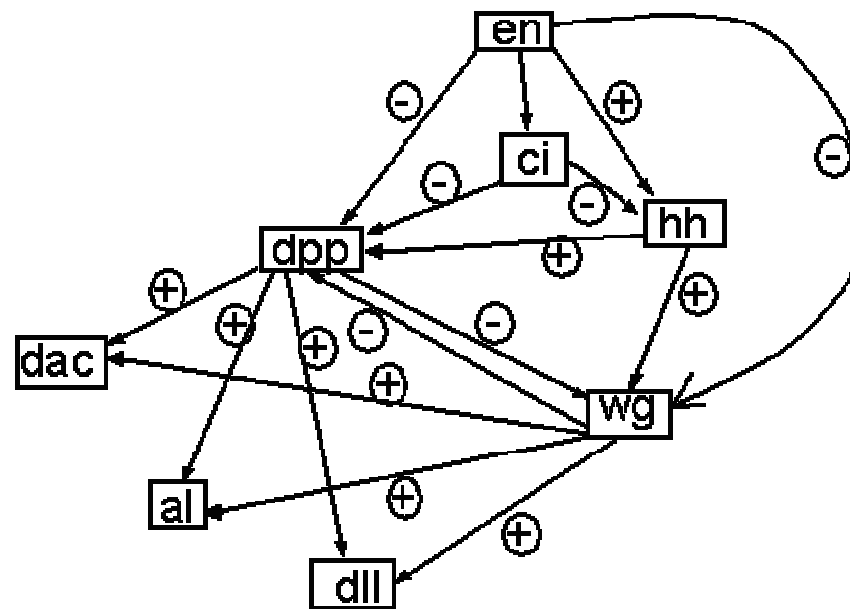


Кассетная активация



Наблюдаемое vs существующее; переключения режимов

Взаимодействия внутри любой сети может быть описано с помощью некоего преобразования. В синхронных сетях значения во всех вершинах меняются одновременно



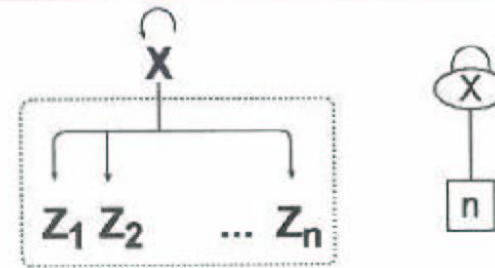
Генная сеть роста лапки
дрозофилы

Разные типы регуляции

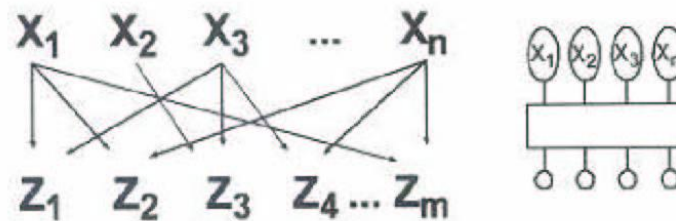
feedforward loop



single input module (SIM)



dense overlapping regulons (DOR)

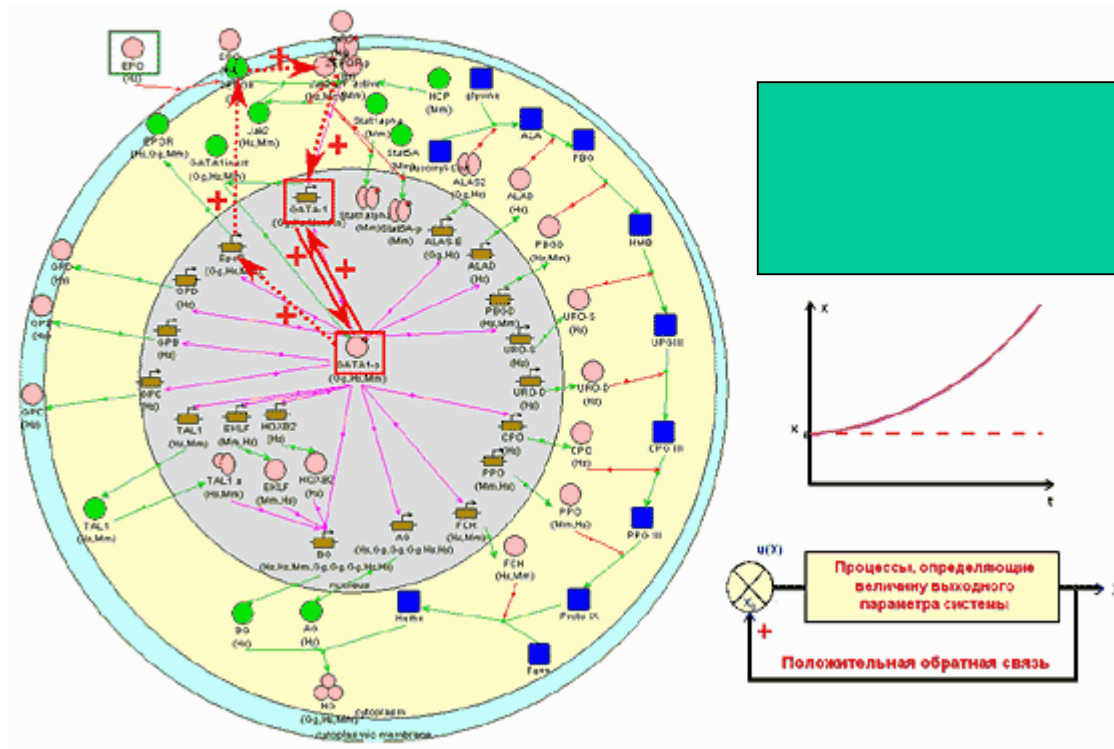


Какие бывают сетевые взаимодействия?

Положительная обратная связь

БАЗА ДАННЫХ GENENET:

ФРАГМЕНТ ГЕННОЙ СЕТИ ДИФФЕРЕНЦИРОВКИ И СОЗРЕВАНИЯ ЭРИТРОЦИТОВ

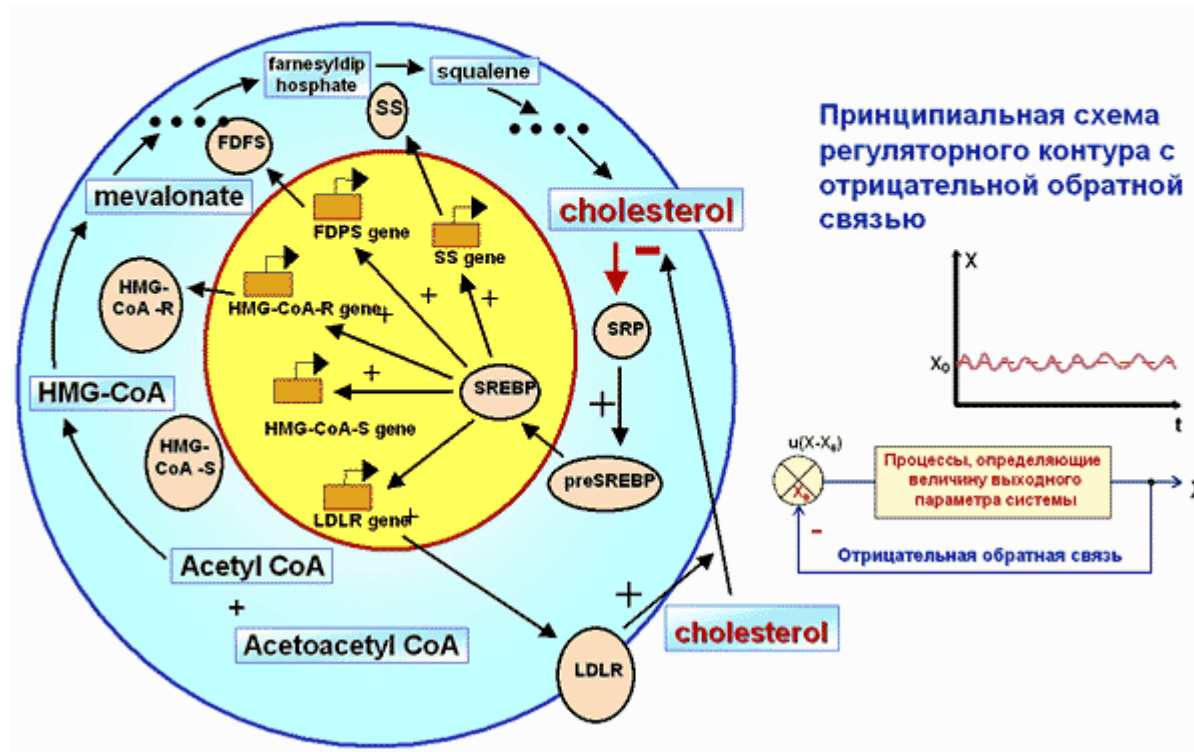


$$\frac{\partial \chi}{\partial t} \propto \chi$$

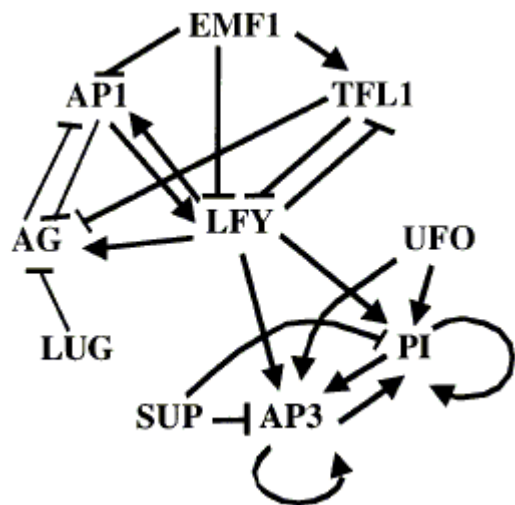
отрицательная обратная связь

БАЗА ДАННЫХ GENENET:

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



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

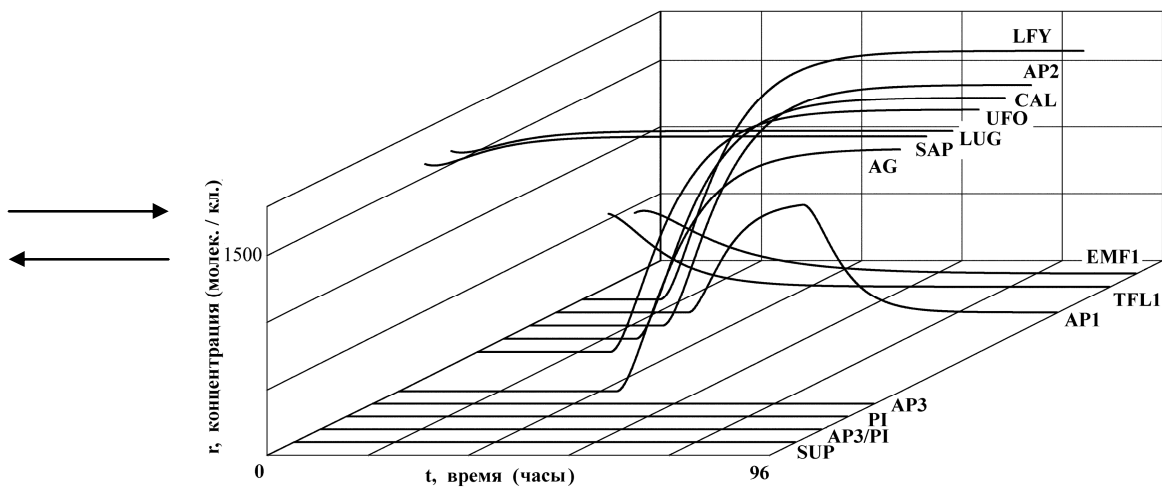
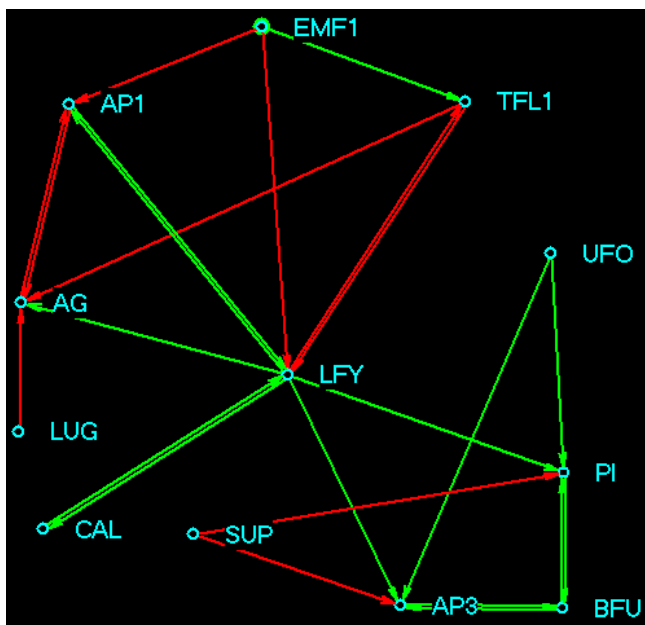


1. Прямая задача для генных сетей

Дана структура и параметры - найти поведение

2. Обратная задача

дано поведение - найти структуру и параметры



Описание преобразования

- Для каждой вершины рассматриваем только входящие дуги
- Рассматриваем активные дуги (если начало дуги 1 дуга активна, иначе дуга неактивна)
- Если нет ни одного активной дуги любого типа, то выходное значение не меняется
- Если есть активная дуга типа «-1» на выходе получаем 0;
- Если нет активных дуг типа «-1» и есть активная дуга типа «1», то на выходе получаем 1

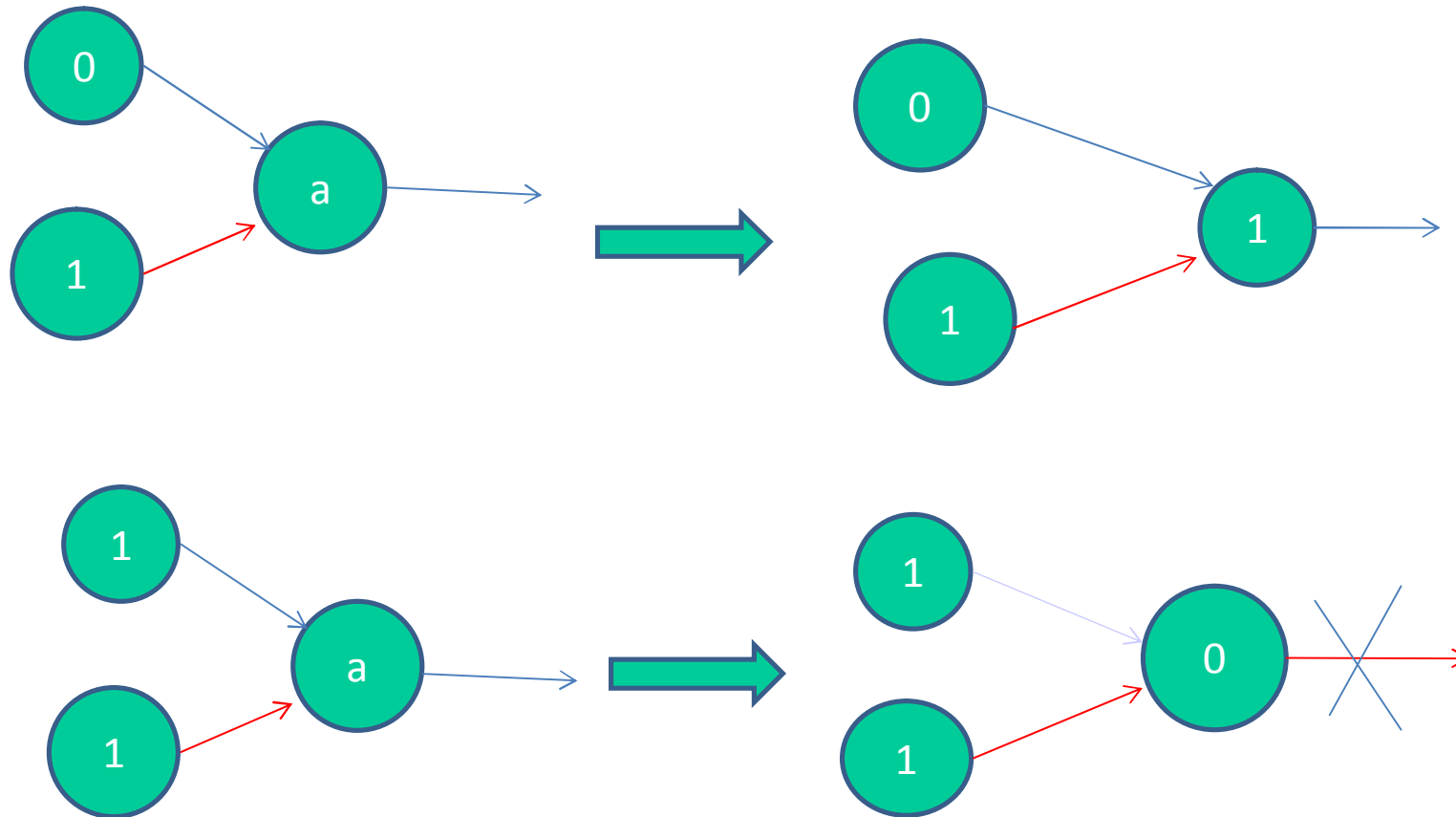
Формальное описание преобразования

- A - $n \times n$ матрица смежности нашего графа, рассматриваемого как взвешенный граф
- V - n булевый вектор
- $|A|$ -матрица, $|A|[i][j]=|A[i][j]|$

Тогда

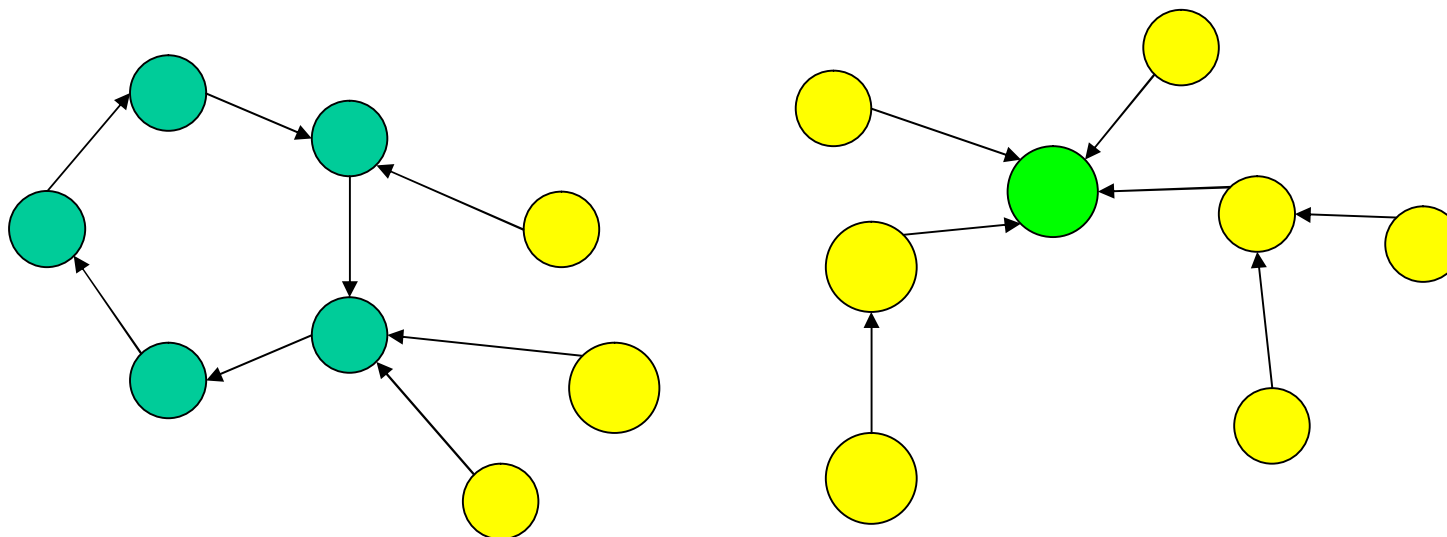
- $F(V[i]) = \begin{cases} 0, & \text{если } (V * A)[i] < (V * |A|)[i] \\ 1, & \text{если } (V * A)[i] = (V * |A|)[i] \neq 0; \\ V[i], & \text{если } (V * A)[i] = (V * |A|)[i] = 0 \end{cases}$

Правила взаимодействия вершин логических сетей



а равно 0 или 1

Конфигурационное пространство сети

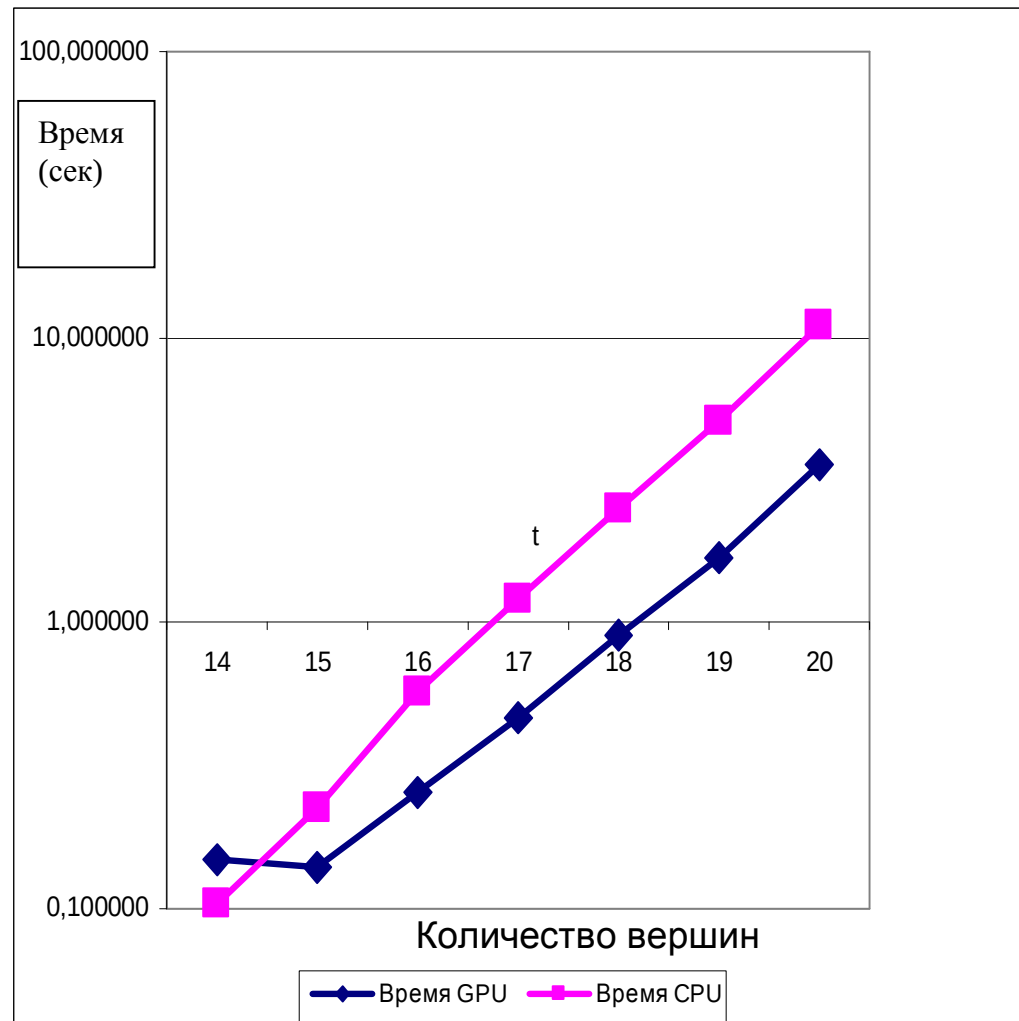


Синим обозначен цикл, желтым области притяжения, зеленым неподвижная точка

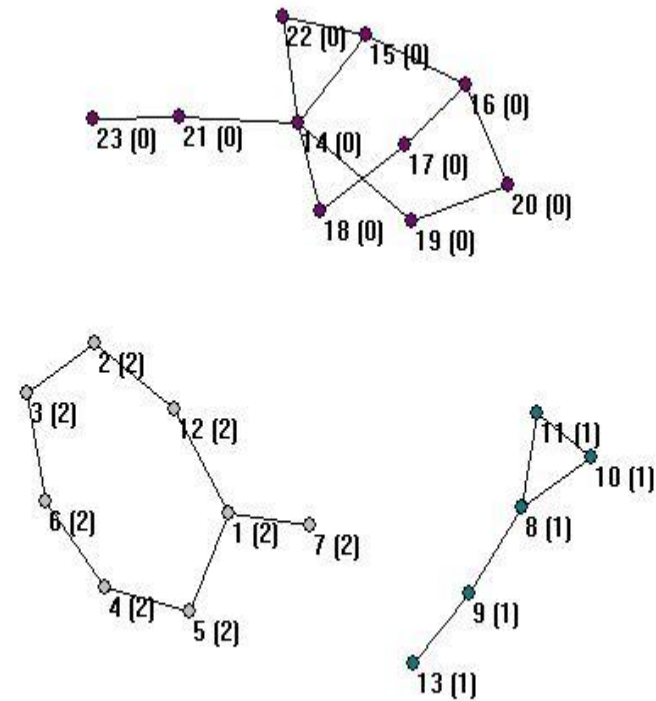
Сравнение времени обработки на CPU и GPU

CPU: Intel 3 GHz

GPU: NVidia GeForce 8400



- 1 "rpoS"
- 2 "RssB-p"
- 3 "RssB"
- 4 "ArcA"
- 5 "ArcA-p"
- 6 "ArcB"
- 7 "H-NS"
- 8 "fnr"
- 9 "YfiA"
- 10 "fnr_o"
- 11 "fnr"
- 12 "RPOs"
- 13 "MMC"
- 14 "OxyR"
- 15 "OxyRr"
- 16 "H2O2"
- 17 "katG"
- 18 "KatG"
- 19 "trxC"
- 20 "TrxC"
- 21 "fur"
- 22 "oxyR"
- 23 "Fur"



Ключевые регуляторы генной сети E.coli
реакции на стресс