

## КОНТАКТНЫЕ ЯВЛЕНИЯ В СЕТЯХ СО СТАТИСТИКОЙ Q – ТИПА

Гаджиев Б. Р., Калинин Е. А., Крюков Ю. А., Прогулова Т. Б.

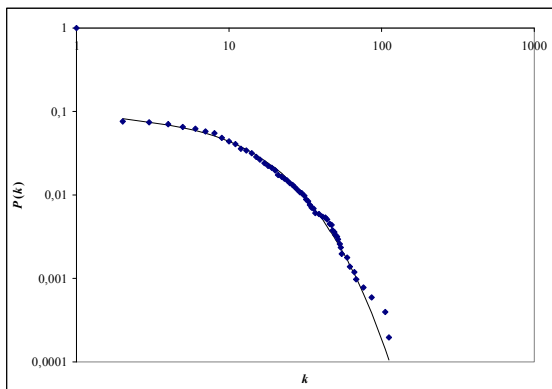
*В данной работе мы изучали компьютерную сеть, состоящую из  $\sim 5 \cdot 10^3$  компьютеров, являющуюся подсетью Интернет. Проведен анализ данных, построены распределения степеней, промежуточной значимости, расстояний между узлами, а также зависимости средней степени ближайших соседей вершины и промежуточной значимости от степени вершин. Предложена модель растущей сети, на основе которой выведена статистика q-типа, адекватно описывающая изучаемую сеть. В соответствии с предложенной моделью имитировались статистически эквивалентные сети, на которых исследованы процессы распространения компьютерных вирусов и оптимальные стратегии лечения*

**Введение.** Для эффективного управления непрерывно растущим трафиком глобального Интернет важно, чтобы генераторы топологии фиксировали его текущую структуру. Это необходимо для планирования и проектирования Интернет. Начиная с открытия Faloutsos в 1999 году известно, что топология Интернет характеризуется степенным распределением степеней [1]. Кроме того, сравнение координат узлов Интернет и карты плотности распределения населения указывает на сильные корреляции между размещением маршрутизаторов и плотностью популяции в экономически развитых областях мира. Анализ размещения маршрутизаторов, доменов и распределения плотности населения методом подсчета покрытий показывает, что каждое из этих трех множеств формирует фрактал с размерностью  $D_f = 1.5 \pm 0.1$  [2]. Топология Интернет изучена на двух различных уровнях — на уровне маршрутизаторов и на уровне автономных систем. В обоих случаях она характеризуется степенным распределением степеней [2]. Показатели распределения степеней для Интернет на уровне автономных систем при размерах сети 3015 – 4389 узлов —  $\gamma = 2.1 - 2.2$ , на уровне маршрутизаторов при размере сети 3888 узлов —  $\gamma = 2.48$ , тогда как на уровне маршрутизаторов при размере сети 150000 узлов —  $\gamma = 2.4$ . Вычисление коэффициента кластеризации показало, что он сравним с коэффициентом кластеризации случайной сети Эрдеша–Реньи и имеет величину  $C_{rand} \approx 0.0001$ .

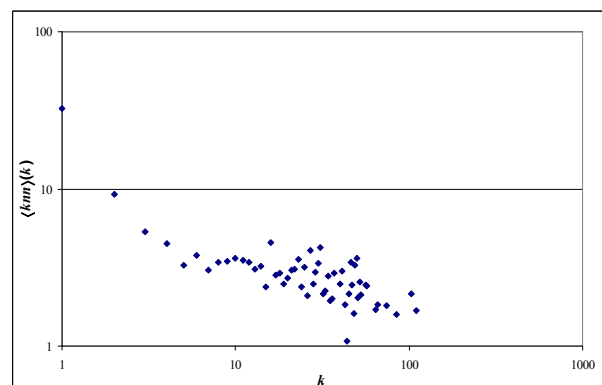
Интернет по определению — это сети физических соединений между компьютерами и другими телекоммуникационными устройствами. Необходимо иметь в виду, что сети растут и развиваются посредством локальных событий, таких как добавление новых узлов или связей, возникновение новых связей между старыми узлами или же пересвязывания [2]. Этот локальный уровень важен для оптимального управления, планирования и проектирования Интернет. Например, при выборе стратегии лечения компьютерных вирусов мы, в самом деле, имеем дело с отдельными узлами сети, а именно, отдельными компьютерами. В связи с этим актуальным представляется детальное исследование подсетей Интернет, состоящих из достаточно большого числа узлов.

В данной работе мы изучали компьютерную сеть, состоящую из  $5 \cdot 10^3$  компьютеров, являющуюся подсетью Интернет. Были проанализированы данные и проведено сравнение с результатами, полученными для глобального Интернет на уровне анализа автономных систем и маршрутизаторов. Далее, предложена модель растущей сети, позволяющая описать результаты обработки данных. Мы также исследовали процесс распространения вирусов в этих сетях (модель *SIS*). Анализировались зависимости доли инфицированных узлов от времени и от скорости распространения вирусов  $\lambda$ . Рассмотрена стратегия уничтожения вирусов, базирующаяся на последовательном лечении узлов с самой высокой степенью (хабы), что приводит к изменению топологии сети и увеличению  $\lambda_c$ .

**Анализ данных.** Мы анализировали данные по городской компьютерной сети, состоящей из  $N \sim 5 \cdot 10^3$  узлов. Максимальная степень вершины  $k_{max} = 112$ . Было построено распределение степеней вершин, представленное на рис. 1. Для выявления наличия и характера корреляций в сети была построена зависимость средней степени ближайших соседей вершины от ее степени  $\langle k_{nn} \rangle(k)$ , представленная на рис. 2.



**Рис. 1.** Кумулятивное распределение степеней, т.е. вероятность того, что вершина имеет степень, большую  $k$

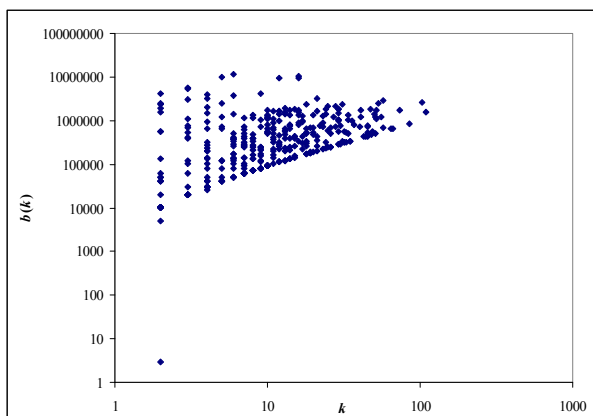


**Рис. 2.** Зависимость  $\langle k_{nn} \rangle(k)$  средней степени ближайших соседей вершины от ее степени

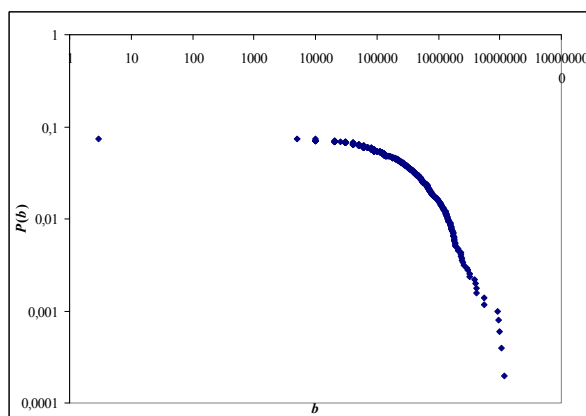
Тот факт, что с ростом  $k$  значения  $\langle k_{nn} \rangle$  уменьшаются, указывает на дисассортативный характер исследуемой сети.

Важной локальной характеристикой является промежуточная значимость узлов сети  $b$  [3]. Значение промежуточной значимости  $b_i^{jl}$  узла  $i$  для вершин  $j$  и  $l$  ( $j \neq l$ ) определяется как доля геодезических между этими вершинами, проходящих через вершину  $i$ . Общая промежуточная значимость вершины  $i$  определяется как  $b_i = \sum_{j \neq l} b_i^{jl}$ . Мы

вычислили значения этой величины для анализируемой сети. На рис. 3 показана зависимость промежуточной значимости от степени вершин, а на рис. 4 — кумулятивное распределение  $P(b)$ , т.е. вероятность того, что вершина имеет значение промежуточной значимости, большее  $b$ . Знание распределения  $P(b)$  необходимо для изучения и управления трафиком.



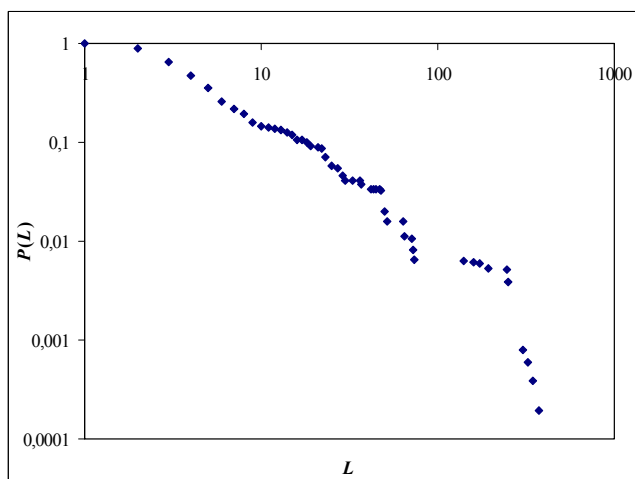
**Рис. 3.** Зависимость  $b(k)$  промежуточной значимости вершин от степени



**Рис. 4.** Кумулятивное распределение промежуточной значимости вершин  $P(b)$

Анализ показывает, что коэффициент кластеризации сети  $C$  приблизительно равен нулю, что свидетельствует о том, что сеть является деревом.

Изучаемая сеть является пространственной в том смысле, что ее узлы имеют определенные координаты в евклидовом пространстве. Мы определяли расстояния между узлами, кумулятивное распределение расстояний приведено на рис. 5.



**Рис. 5.** Кумулятивное распределение расстояний между узлами (в относительных единицах)

**Модель.** Предлагаемая модель основывается на том, что процесс формирования подсети интернет осуществляется посредством следующих трех типов локальных событий: добавление новых узлов, возникновение новых связей между старыми узлами или же пересвязывания ребер [2]. Ниже мы рассмотрим непрерывную модель.

(i) С вероятностью  $p$  между уже существующими узлами добавляется  $m$  новых связей. Одна из вершин для каждого ребра при этом выбирается случайно, а другая — в соответствии с принципом предпочтительного присоединения. Таким образом,

$$\left(\frac{\partial k_i}{\partial t}\right)_{(i)} = pA \frac{1}{N} + pA \frac{k_i + 1}{\sum_j (k_j + 1)} \quad (1)$$

где  $N$  — размер системы. Так как изменение суммарной степени после очередного шага есть  $\Delta k = 2m$ , то  $A = m$ .

(ii) С вероятностью  $q$  происходит пересвязывание  $m$  ребер. Уравнение для изменения степени имеет вид:

$$\left(\frac{\partial k_i}{\partial t}\right)_{(ii)} = -qB \frac{1}{N} + qB \frac{k_i + 1}{\sum_j (k_j + 1)} \quad (2)$$

Здесь первый член описывает уменьшение степени вершины, от которой ребро было отсоединено, а второй член — увеличение степени вершины, к которой ребро присоединили. Эта вторая вершина выбирается в соответствии с принципом предпочтительного присоединения. Суммарная степень не изменяется, и, можно показать,  $B = m$ .

(iii) С вероятностью  $1 - p - q$  в сеть поступает новая вершина, соединяющаяся  $m$  ребрами с уже присутствующими в сети различными узлами в соответствии с принципом предпочтительного присоединения. Таким образом,

$$\left(\frac{\partial k_i}{\partial t}\right)_{(iii)} = (1 - p - q)C \frac{k_i + 1}{\sum_j (k_j + 1)}. \quad (3)$$

Число связей, соединяющих новый узел с узлами, уже присутствующими в системе, равно  $m$ , и, таким образом,  $C = m$ .

Суммируя вклад процессов (1), (2) и (3), получаем

$$\frac{\partial k_i}{\partial t} = \left(\frac{\partial k_i}{\partial t}\right)_{(i)} + \left(\frac{\partial k_i}{\partial t}\right)_{(ii)} + \left(\frac{\partial k_i}{\partial t}\right)_{(iii)} = (p - q)m \frac{1}{N} + (1 - p - q) \frac{k_i + 1}{\sum_j (k_j + 1)} \quad (4)$$

Поскольку размер системы  $N$  и суммарная степень  $\sum_j k_j$  изменяются со временем как  $N(t) = m_0 + (1 - p - q)t$  и  $\sum_j k_j = (1 - q)2mt$ , для больших  $t$  мы можем пренебречь  $m_0$  по сравнению с членами, линейно увеличивающимися со временем. Тогда  $N(t) = (1 - p - q)t$  и  $\sum_j (k_j + 1) = (1 - q)2mt + N = (1 - q)2mt + (1 - q - p)t$ , и уравнение (4) принимает вид:

$$\frac{\partial k_i}{\partial t} = (p - q)m \frac{1}{(1 - p - q)t} + (1 - p - q)m \frac{k_i + 1}{(1 - q)2mt + (1 - q - p)t}. \quad (5)$$

Введем обозначения

$$A(p, q, m) = (p - q) \left( \frac{2m(1 - q)}{1 - p - q} + 1 \right) \text{ и } B(p, q, m) = \frac{2m(1 - q) + (1 - p - q)}{m}.$$

Тогда уравнение (5) можно записать в виде

$$\frac{\partial k_i}{\partial t} = [A(p, q, m) + 1 + k_i] \frac{1}{B(p, q, m)t} \quad (6)$$

Решение уравнения (6) имеет вид  $k_i(t) = [A(p, q, m) + 1 + m \left( \frac{t}{t_i} \right)^{\frac{1}{B(p, q, m)}} - A(p, q, m) - 1$ .

Вероятность, что узел имеет степень  $k_i(t)$  меньшую, чем  $k$ ,  $P[k_i(t) < k]$ , можно записать как  $P[k_i(t) < k] = P[t_i > C(p, q, m)t]$ , где введено обозначение

$$C(p, q, m) = \frac{(A(p, q, m) + m + 1)^{B(p, q, m)}}{(k + m + 1)^{B(p, q, m)}}.$$

Так как  $t_i$  должно удовлетворять условию  $0 \leq t_i \leq t$ , мы можем различать следующие случаи:

(i) Если  $C(p, q, m) > 1$ , то  $P[k_i(t) < k] = 0$ .

(ii) Если  $0 < C(p, q, m) < 1$ , то распределение степеней  $P(k)$  может быть определено аналитически.

Так добавление связей, пересвязывание и рост происходят однородно во времени,  $P_i(t_i) = \frac{1}{m_0 + t}$ , то  $P[k_i(t) < k] = P[t_i > C(p, q, m)t] = 1 - P[k_i(t) > k] = 1 - \frac{C(p, q, m)t}{m_0 + t}$ , откуда, используя

определение  $P(k, t) = \frac{\partial P[k_i(t) < k]}{\partial k}$ , для распределения степеней получаем

$$P(k, t) = D(p, q, m) \frac{t}{m_0 + t} (k + A(p, q, m) + 1)^{-1 - B(p, q, m)},$$

где  $D(p, q, m) = B(p, q, m) [m + A(p, q, m) + 1]^{B(p, q, m)}$ . Таким образом, распределение степеней имеет обобщенную степенную форму. Распределение степеней — основная топологическая характеристика сети, которая определяет вероятность того, что случайно выбранная вершина в сети в момент времени  $t$  имеет степень  $k$ , или, эквивалентно, определяет в момент времени  $t$  долю узлов, имеющих степень  $k$  [3].

В пределе  $t \rightarrow \infty$  мы получаем стационарное распределение степеней:

$$P_{st}(k) = D(p, q, m) (k + A(p, q, m) + 1)^{-1 - B(p, q, m)}.$$

Вводя обозначения  $\kappa(p, q, m) = A(p, q, m) + 1$  и  $\gamma(p, q, m) = B(p, q, m) + 1$ , получаем

$$P_{st}(k) = D(p, q, m) (\kappa(p, q, m) + k)^{-\gamma(p, q, m)}.$$

Последнее выражение можно записать в виде

$$P_{st}(k) = D(p, q, m) (\kappa(p, q, m))^{-\gamma(p, q, m)} \left( 1 + \frac{1}{\kappa(p, q, m)} k \right)^{-\gamma(p, q, m)}.$$

Пусть  $\frac{1}{\kappa(p, q, m)} = -(1 - q) \frac{1}{k_0}$  и  $\frac{1}{1 - q} = -\gamma(p, q, m)$ . Тогда  $k_0 = \frac{\kappa(p, q, m)}{\gamma(p, q, m)}$  и

$$P_{st}(k) = D(p, q, m) (\kappa(p, q, m))^{-\gamma(p, q, m)} \left( 1 - (1-q) \frac{k}{k_0} \right)^{\frac{1}{1-q}}.$$

Используя условие нормировки вероятности  $P_{st}(k)$ , можно записать

$$P_{st}(k) = \frac{1}{Z} \left( 1 - (1-q) \frac{k}{k_0} \right)^{\frac{1}{1-q}}, \text{ где } Z = \int_0^{\infty} \left( 1 - (1-q) \frac{k}{k_0} \right)^{\frac{1}{1-q}} dk. \quad (7)$$

Таким образом, мы приходим к распределению Цаллиса [4], которое при  $q \rightarrow 1$  имеет вид экспоненциального распределения, тогда как при больших  $k$  имеет вид степенного распределения.

**Неэкстенсивная статистика. Вывод распределения степеней и распределения длин.** Мы можем определить функцию распределения степеней, используя принцип максимума энтропии. Так как мы имеем дело со сложными системами, можно использовать неэкстенсивную статистическую механику. В рамках этого подхода энтропия  $S_q$  представляется следующим образом [4]:

$$S_q = \frac{\sum p_k^q - 1}{1-q}, \quad (8)$$

где  $q$  — индекс энтропии распределения вероятностей  $p_k$  состояний  $k$ . Естественными ограничениями для максимизации энтропии (8) являются

$$\sum_k p_k = 1, \quad (9)$$

что соответствует условию нормировки вероятностей,

$$\sum_k P(k)k = \mu, \quad (10)$$

где  $\mu$  — первый момент распределения  $P(k)$  и

$$P(k) = \frac{p_k^q}{c_q}, \quad c(q) = \sum_k p_k^q, \quad (11)$$

$P(k)$  — распределение вероятностей степеней свободы.

Из вариационной проблемы для (8) с учетом ограничений (9) и (10) получаем

$$\frac{\delta}{\delta p_k} \left( \frac{\sum_{k'} p_{k'}^q - 1}{1-q} - \beta \sum_{k'} P(k')k' - \lambda \sum_{k'} p_{k'} \right) = 0,$$

где  $\alpha$  и  $\beta$  — параметры Лагранжа. Решение имеет вид

$$p_k = (g)^{\frac{1}{1-q}} \left( \frac{\lambda(1-q)}{q} \right)^{\frac{1}{q-1}} \left[ 1 - \frac{(1-q)}{c_q} \beta k \right]^{\frac{1}{1-q}}.$$

С учетом (9)  $\lambda$  определяется уравнением

$$g^{\frac{1}{1-q}} \left( \frac{\lambda(1-q)}{q} \right)^{\frac{1}{q-1}} \sum_k (1-(1-q) \frac{\beta k}{c_q})^{\frac{1}{1-q}} = 1$$

Обозначим  $Z = \sum_k (1-(1-q) \frac{\beta k}{c_q})^{\frac{1}{1-q}}$ , тогда  $g^{\frac{1}{1-q}} \left( \frac{\lambda(1-q)}{q} \right)^{\frac{1}{q-1}} = Z^{-1}$  и, таким образом

$$p_k = \frac{1}{Z} \left( 1 - (1-q) \frac{\beta}{c_q} k \right)^{\frac{1}{1-q}}$$

Используя определение (11), получаем

$$P(k) = \frac{(1 - \xi(1-q)k)^{\frac{q}{1-q}}}{\sum_k (1 - \xi(1-q)k)^{\frac{q}{1-q}}}$$

Полученное выражение представляет собой распределение Цаллиса.

Аналогично можно определить распределение длин ребер. В этом случае кроме условия нормировки  $\int_0^{\infty} P(x) dx = 1$  мы вводим ограничение  $\int_0^{\infty} (x - x_0) P(x) dx = 0$ .

Неэкстенсивная энтропия определяется как

$$S_q = \frac{\int_0^{\infty} p^q(x) dx - 1}{1 - q}.$$

Анализ, основанный на методе Лагранжа, приводит к следующему результату

$$P(x) = \frac{(1 - \xi(1-q)x)^{\frac{q}{1-q}}}{\int_0^{\infty} dx (1 - \xi(1-q)x)^{\frac{q}{1-q}}}. \quad (12)$$

Это распределение описывает распределение расстояний между узлами сети и представляет собой распределение Цаллиса для расстояний.

**Распространение вирусов и стратегия лечения.** В компьютерных системах без автоматического обновления антивирусных программ распространение вирусов в сетях с заданной топологией описывается моделью *SIS*, в которой предполагается, что узлы в сети могут находиться в одном из двух состояний: восприимчивом (*S*) и инфицированном (*I*) [5]. В этой модели восприимчивые узлы инфицируются, переходят в инфицированное состояние и снова становятся восприимчивыми спустя некоторый промежуток времени, по истечении которого они становятся здоровыми. Затем, они снова могут подвергнуться инфицированию. Таким образом, процесс распространения вирусов описывается реакцией  $S(i) + I(j) \xrightarrow{\lambda} I(i) + I(j)$ ,  $I(i) \xrightarrow{\mu} S(i)$ , где  $\lambda$  и  $\mu$  — вероятности соответствующих переходов. Для конкретной сети процесс распространения определяется ее топологией и вероятностью  $\lambda$  перехода узла из состояния *S* в состояние *I*

(то есть, здоровый узел в сети при контакте с зараженным узлом становится инфицированным). Можно показать, что существует пороговое значение вероятности перехода  $\lambda_c$ , такое, что при значениях  $\lambda < \lambda_c$  за конечное время доля инфицированных узлов  $\rho$  в сети становится равной нулю — распространение вирусов затухает, при  $\lambda > \lambda_c$  доля инфицированных узлов  $\rho$  будет больше нуля, т.е. вирусы в сети распространяются неограниченно долго.

Пороговое значение параметра  $\lambda$  определяется из условия

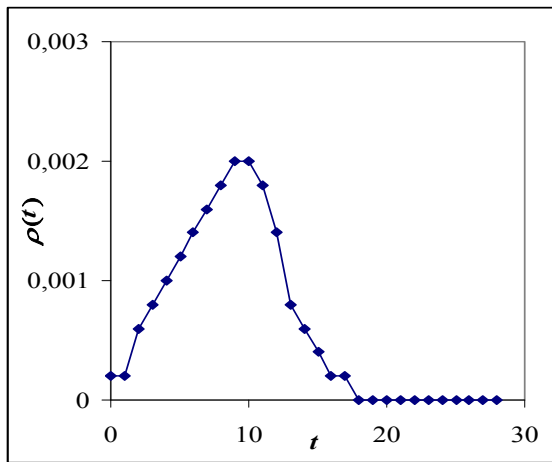
$$\frac{d}{d\Theta} \left( \frac{1}{\langle k \rangle} \sum_k k P(k) \frac{k\lambda\Theta(\lambda)}{1+k\lambda\Theta(\lambda)} \right)_{\Theta=0} = 1, \quad (13)$$

где  $\langle k \rangle$  — средняя степень узлов, а  $\Theta(k)$  определяется из решения уравнения [6]

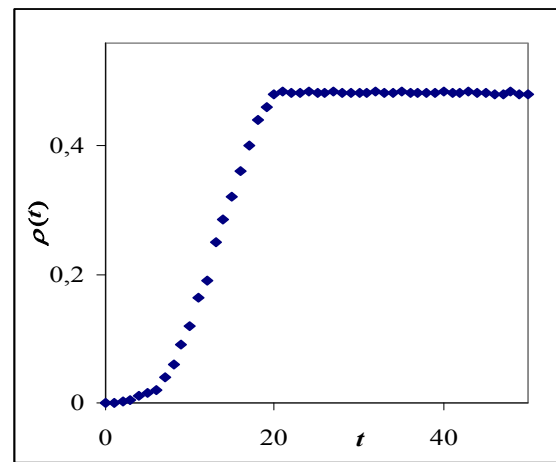
$$\Theta(\lambda) = \frac{1}{\langle k \rangle} \sum_k k P(k) \frac{k\lambda\Theta(\lambda)}{1+k\lambda\Theta(\lambda)}.$$

В нашем случае  $P(k)$  — распределение Цаллиса. Из (13) находим, что  $\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle}$ .

Мы генерировали сети, статистически эквивалентные реальной исследуемой, с параметрами  $p$ ,  $q$  и  $t$ , выбранными таким образом, чтобы топология сгенерированной сети совпадала с топологией анализируемой сети. Далее, мы изучали в сгенерированных сетях процесс распространения вирусов в рамках модели *SIS*. Зависимость от времени плотности инфицированных узлов  $\rho(t)$  ниже порога  $\lambda < \lambda_c$  приведена на рисунке 6, а выше порога  $\lambda > \lambda_c$  на рисунке 7.



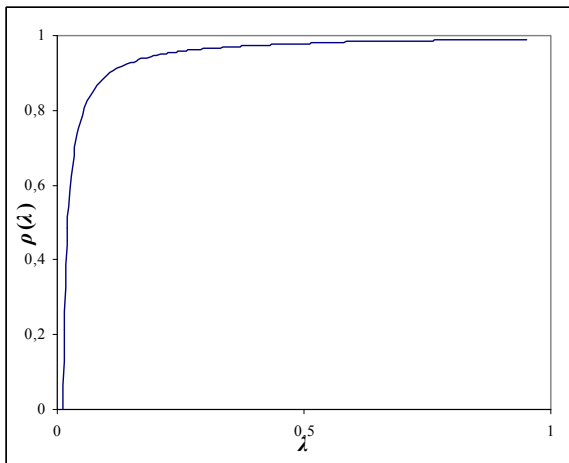
**Рис. 6.** Зависимость плотности инфицированных узлов от времени при  $\lambda < \lambda_c$



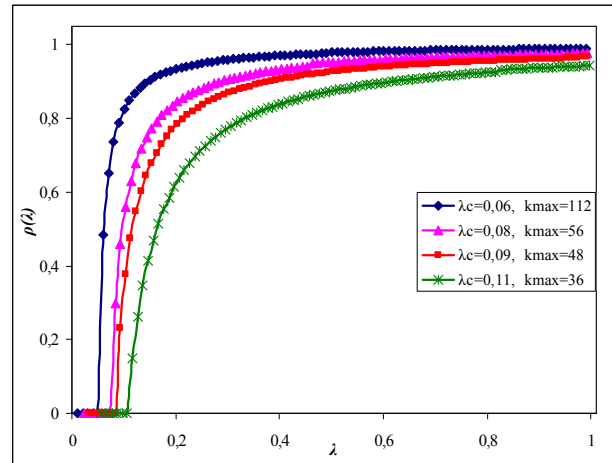
**Рис. 7.** Зависимость плотности инфицированных узлов от времени при  $\lambda > \lambda_c$

Зависимость от параметра  $\lambda$  плотности инфицированных узлов  $\rho(\lambda)$  приведена на рисунке 8.





**Рис. 8.** Зависимость плотности инфицированных узлов  $\rho(\lambda)$  от параметра  $\lambda$



**Рис. 9.** Зависимость плотности инфицированных узлов  $\rho(\lambda)$  от параметра  $\lambda$  при удалении из сети вершин со степенью большей  $k_{max}$

Рассмотрена стратегия уничтожения вирусов, базирующаяся на последовательном лечении узлов с самой высокой степенью (хабы), что приводит к изменению топологии сети и увеличению  $\lambda_c$ . На рисунке 9 представлены зависимости плотности инфицированных узлов от параметра  $\lambda$  для исходной сети и для сетей, получаемых при последовательном удалении из сети (лечения) зараженных вершин с максимальной степенью.

**Обсуждение результатов.** Наш анализ городской сети, которая является подсетью Интернет, показывает, что ее распределение степеней описывается распределением Цаллиса. Используя метод максимального правдоподобия, из подгонки мы определили значение параметра неэкстенсивности:  $q = 1.1$ . Как уже было отмечено, при больших степенях распределение Цаллиса ведет себя как  $p(k) \sim k^{-\gamma}$ . В нашем случае определенное по хвосту распределения значение  $\gamma \cong 2.8$ . Таким образом, полученное значение не совпадает с показателями, найденными для глобального Интернет на уровне автономных систем и на уровне маршрутизаторов. Это свидетельствует о мультифрактальной природе сети Интернет.

Результаты подгонки показывают также, что определяющим является для исследуемой сети не процесс связывания узлов, уже существующих в сети, и не процесс пересвязывания, а именно, процесс роста. Это подтверждается также полученными результатами по изучению характера корреляций степеней вершин (зависимость  $\langle k_{nn} \rangle(k)$  показывает, что сеть дисассортативна) и близостью к нулю среднего значения коэффициента кластеризации. Мы также анализировали распределение длин и получили, что  $p(L) \sim L^{-\alpha}$ , где  $\alpha = 1.2$ .

Для сети со степенным распределением порог эпидемии равен нулю [6]. Поэтому масштабно-инвариантные сети уязвимы при распространении эпидемий компьютерных вирусов. Конечность сети приводит к конечному порогу эпидемии. Однако эта величина все равно остается достаточно маленькой, что мы наблюдаем и для анализируе-

мой сети:  $\lambda_c \cong 0,06$ . Поэтому стратегия остановки эпидемии заключается в последовательном лечении вершин наибольшей степени, в результате которого топология сети меняется, что приводит к росту значения  $\lambda_c$ .

#### СПИСОК ЛИТЕРАТУРЫ

1. Faloutsos M., Faloutsos P., Faloutsos C. On Power-Law Relationships of the Internet Topology // ACM SIG-COMM '99 Comput. Commun Rev. 1999. Vol. 29. P. 251.
2. Yook S.-H. Jeong H. and Barabási A.-L. Modeling the Internet's large-scale topology // Proceedings of the Nat'l Academy of Sciences. 2000. Vol. 99. P. 13382-13386.
3. Albert R., Barabasi A.-L. Statistical Mechanics of Complex Networks // Rev. Mod. Phys. 2002. Vol. 74, №1. P. 43–97.
4. Tsallis C. Nonextensive statistics: theoretical, experimental and computational evidences and connections// Braz. J. Phys. 1999. Vol. 29. P. 1-35.
5. Pastor-Satorras R., Vespignani A. Evolution and Structure of the Internet: A Statistical Physics Approach. — Cambridge: Cambridge University Press, 2004. 267 p.
6. Dezzo Z. and Barabási A.-L. Halting viruses in scale-free networks // Phys. Rev. E. 2002. Vol. 65. P. 055103.

#### THE CONTACT PHENOMENA IN NETWORKS WITH Q – TYPE STATISTICS

**Gadjiev B. R., Kalinkina E. A., Kryukov Yu. A., Progulova T. B.**

*We investigate the computer network consisting of  $\sim 5 \cdot 10^3$  computers which is a subnet of the Internet. On the base of data analysis we determined distribution functions of degree, of betweenness centrality and of distances between nodes. Dependences of an average degree of the nearest neighbours vertices and betweenness centrality on a degree of nodes are constructed. The model of a growing network is introduced and the q-type statistics is derived on its basis. It is shown that the results of this model adequately describe the studied network. According to the offered model we simulated statistically equivalent networks on which computer viruses spreading processes are investigated. We find that the more biased a policy is towards the hubs, the more chance it has to bring the epidemic threshold above the virus' spreading rate*