

**Ордена Ленина
ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ
имени М.В.Келдыша
Российской академии наук**

Н.А. Митин, А.В. Подлазов, Д.П. Щетинина

**Исследование сетевых свойств
Живого журнала**

Москва — 2013

Н.А. Митин, А.В. Подлазов, Д.П. Щетинина

Исследование сетевых свойств Живого журнала

Исследуется структура связей в социальной сети Живого журнала (ЖЖ). Проанализированы ее основные количественные характеристики: распределение вершин по числу соседей, корреляция при образовании связей и кластеризация вторичных связей. Выявлен масштабно-инвариантный характер этих зависимостей в широких диапазонах значений. Продемонстрирована ассортативность сети.

Рассмотрено распределение кратчайших путей, связывающих пары пользователей, которое имеет отчетливо выраженный пик на шести звеньях. Тем самым подтверждается концепция шести степеней разделения С. Милгрэма. Близкое взаиморасположение большинства пользователей наряду с масштабно-инвариантным устройством сети ЖЖ позволяет отнести его к малым мирам.

Отдельно исследована устойчивость ЖЖ как среды распространения широковещательных сообщений по отношению к направленным атакам и случайным отказам.

Ключевые слова: малые миры, социальные сети, масштабная инвариантность, степенные распределения, вершинная корреляция, ассортативность, кластеризация вторичных связей, шесть степеней разделения, устойчивость к направленным атакам и случайным отказам.

N.A. Mitin, A.V. Podlazov, D.P. Shchetinina

The Study of Live Journal Network Properties

We investigate the connection structure in social network Live Journal (LJ) using main quantitative characteristics of its network such as the degree distribution, correlations, and clustering of secondary connections. These dependences are found to be scale invariant in a wide variety of values. The network is the assortative one.

The shortest paths distribution between users has the sharp peak on six links. This confirms Milgram's concept of six degrees of separation. Close location of most LJ-users along with its scale invariant structure allows treating it as small world.

In addition, we study the resistance of LJ as environment of broadcast messages spreading with regard to targeted attacks and random failures.

Key words: small-worlds, social networks, scale invariance, power laws, vertex correlation, assortativity, the clustering of secondary connections, six degrees of separation, the resistance to targeted attacks and random failures.

Работа выполнена при поддержке РФФИ (проект №12-06-00402).

ВВЕДЕНИЕ

В последние годы возрос интерес к исследованию социальных сетей.

Впервые идею использовать теорию сложных систем для исследования социальных сетей высказал профессор Гарвардского университета Стивен Милгрэм в 1967 году [1]. В своем эксперименте, заключавшемся в определении «расстояния» между любыми двумя людьми в Соединенных Штатах Америки, он проверил гипотезу *шести степеней разделения*. В соответствии с ней любые два человека находятся в среднем в шести связях друг от друга (связь здесь означает знакомство, иногда говорят о шести рукопожатиях). С. Милгрэм показал, что люди не только взаимосвязаны, но и живут в *малом*, или *тесном*, мире, что общество представляет очень плотную сеть знакомств.

Развитие коммуникационных систем, таких как телефонная сеть и Интернет, способствовало появлению новых форм социальных контактов. Примерами сетей, порожденных такими контактами, являются сети телефонных звонков, сеть обмена письмами по электронной почте и социальные сети, такие как «ВКонтакте», «Одноклассники», «Живой журнал», «twitter», «Facebook» и др.

Мы рассматриваем одну из таких социальных сетей – Живой журнал (ЖЖ). ЖЖ представляет собой программную поддержку общения через Интернет некоторого множества людей, которые должны быть зарегистрированы в сети ЖЖ (зарегистрированные пользователи) и множества направленных связей, которые участник ЖЖ определяет как связи со своими друзьями, т.е. каждый участник ЖЖ сам определяет круг своих друзей. Установленные связи могут быть односторонними. Организация этой социальной сети позволяет получать некоторый набор данных об ее участниках и их связях. Эти данные были собраны на сайте <http://www.livejournal.com> по состоянию на апрель 2011 года. Предметом исследования являются статистические характеристики множества участников ЖЖ и их связей.

Вначале работы дается описание сети Живого журнала и рассматривается ее структура. Затем анализируется топология и корреляции в данной сети, исследуется структура вторичных связей на основе коэффициента кластеризации. После этого изучается структура минимальных расстояний между пользователями. Наконец, анализируется устойчивость ЖЖ к направленным атакам на сеть и случайным отказам вершин. В заключении приводятся основные результаты исследованной сети ЖЖ.

СТРУКТУРА ЖИВОГО ЖУРНАЛА

Живой журнал был создан американским программистом Брэдом Фитцпатриком, который зарегистрировал себя впервые под ником «brad» 18 марта 1999 года. За годы существования Живой журнал разросся во всемирную социальную сеть. 25 ноября 2009 года Живой журнал стал лауреатом конкурса «Премия Рунета» в номинации «Государство и общество» и, согласно стати-

стике в Рунете, «Живой журнал» имеет значительную посещаемость, функционируя в роли популярной социальной сети.

Живой журнал позволяет вести ежедневные онлайн-дневники, составлять списки своих друзей, дневники которых интересны для чтения, узнавать новости друзей и оставлять комментарии по поводу прочитанного.

На момент скачивания данных в Живом журнале было зарегистрировано свыше 7,8 млн активных пользователей (имеющих хотя бы одну связь с другими пользователями) с 284 млн связями (количество исходящих и входящих связей одинаково).

В структуре ЖЖ имеются *потерянные* связи, в отношении которых запросы данных о пользователях, находящихся на разных концах связей, дают разные сведения. Т.е. пользователь на одном конце связи «думает», что она есть, а на другом конце – что ее нет. Как правило, потерянные связи возникают у пользователей с очень большим числом связей. При анализе структуры сети мы реконструировали потерянные связи, полагая, что если хотя бы один пользователь «знает» о связи, то она есть. Общее количество потерянных связей составляет порядка 1 млн.

Если представлять ЖЖ в виде ориентированного графа, то пользователи являются его вершинами, а исходящие ребра указывают на друзей данного пользователя. Граф описывается матрицей смежности $V = \|v_{ij}\|$ размера N , в которой каждый столбец и строка представляют вершины графа. При этом $v_{ij} = 1$, если вершина j связана с вершиной i , и $v_{ij} = 0$ в противном случае. При этом симметричная матрица $W = \|w_{ij}\|$ с элементами $w_{ij} = v_{ij}v_{ji}$ описывает структуру взаимных связей.

1. Топология сети

Современные подходы статистической физики позволили разработать последовательные методы описания топологии и эволюции сложных сетей.

Основной характеристикой вершины i является ее *степень* k_i – количество смежных с ней вершин. Поскольку рассматриваемый граф – направленный и матрица смежности является несимметричной, следует отдельно рассматривать входящие и исходящие связи. Соответствующие им значения характеристик $k_i^{\text{in}} = \sum_{j=1}^N v_{ji}$ и $k_i^{\text{out}} = \sum_{j=1}^N v_{ij}$ будем помечать верхними индексами, а значения для двунаправленных связей $k_i = \sum_{j=1}^N w_{ij}$ будем писать без индекса. Так же мы будем опускать индекс в том случае, когда речь идет об общих формулах, в равной степени применимых к любому типу связей.

Применительно к описанию Живого журнала будем считать направлением связей направление интереса пользователей, противоположное направлению распространения информации. При этом *входящая степень* k_i^{in} вершины определяет число пользователей, считающих пользователя i своим другом и, соот-

ответственно, читающих его посты в своих френд-лентах. Аналогичным образом *исходящая степень* вершины k_i^{out} – число пользователей, которых сам пользователь i считает своими друзьями, получая их посты в свою френд-ленту. Наконец, *двунаправленная (взаимная) степень* k_i выражает число пользователей, испытывающих взаимный интерес с пользователем i (т.е. они читают дневники друг друга).

В отсутствие заданных извне характерных значений вероятность того, что вершина имеет степень, равную k , описывается масштабно-инвариантным законом вида

$$u(k) \sim k^{-(1+\alpha)}.$$

Рис. 1, на котором приведено распределение пользователей ЖЖ по числу связей определенного типа, позволяет убедиться, что данная формула применима к нему лишь в определенном диапазоне значений – *промежуточной асимптотике*. Ее левая граница лежит на уровне $k \sim 100$. Скорее всего, этот

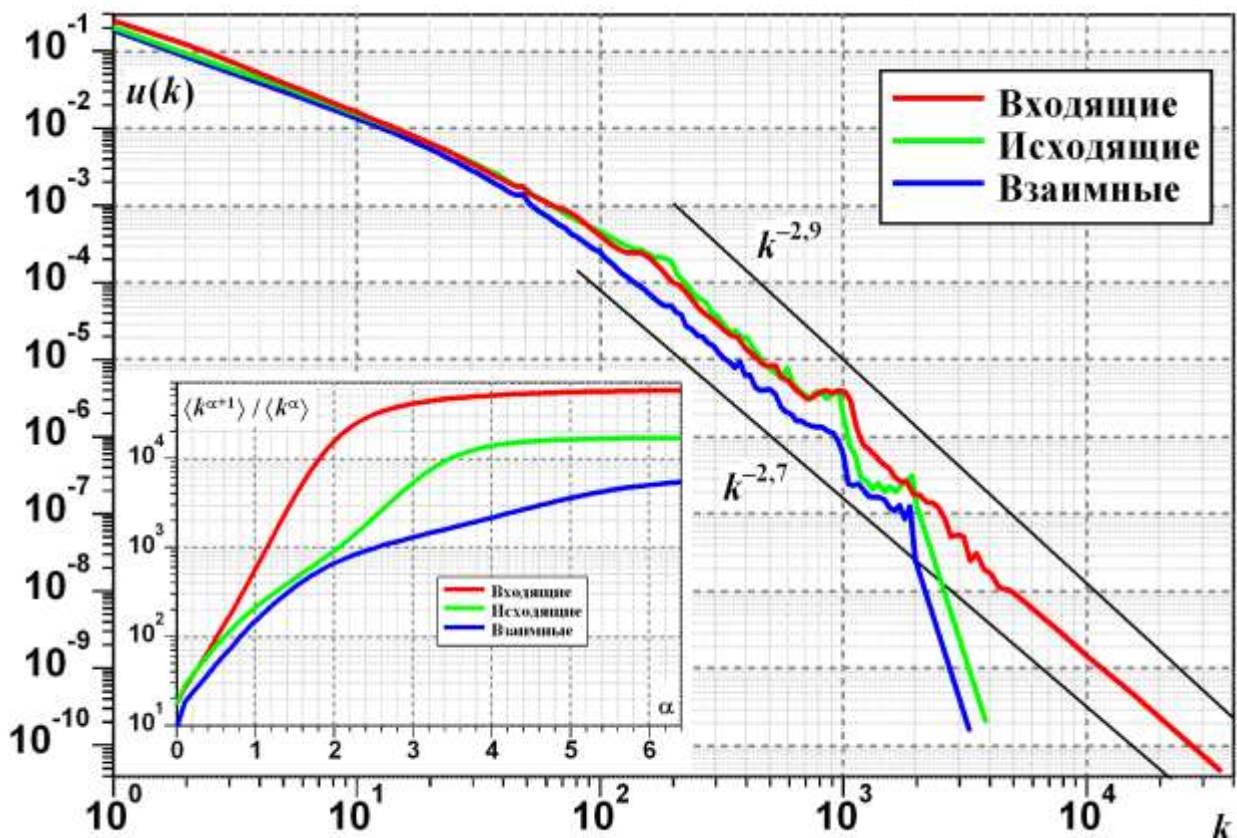


Рис. 1. Распределение пользователей по числу связей

Плотность описывается степенным законом с показателем $\alpha \approx 1,7 \div 1,9$ в промежуточной асимптотике, ширина которой составляет примерно один порядок для исходящих и взаимных связей и чуть более двух – для входящих.

На врезке – отношение последовательных моментов распределения. Переход графиков от возрастания к насыщению соответствует значению α , совпадающему с показателем распределения, а горизонтальная асимптота – максимальным зарегистрированным значениям k .

порог отделяет активных блогеров от новичков или пользователей, которые уже забросили свои дневники. В любом случае можно констатировать, что активность пользователей социальных сетей является в определенном смысле пороговой. Блогеры или читатели блогов, не накопившие определенного багажа сетевых контактов, подчиняются не тем закономерностям, что всегда были у сети.

Правая граница области промежуточной асимптотики для исходящих, а с ними и для взаимных связей имеет значение $k \sim 2\,000$. По всей видимости, при большем количестве друзей человек уже не в состоянии помнить, кто они такие. Тем более невозможным становится даже беглый просмотр френд-ленты, в которую собираются потоки информации из слишком большого числа источников. А вот для входящих связей столь жестких ограничений нет: высокий интерес к блогу не мешает его автору и даже может становиться источником дохода. Поэтому правая граница промежуточной асимптотики для входящих связей на имеющихся данных не регистрируется. Тем не менее, нельзя исключать ее существования, т.к. многие пользователи предпочитают заходить в дневники популярных блогеров напрямую. Если зафрендить их, то идущий от них большой поток постов забьет френд-ленту до такой степени, что в ней будут теряться посты прочих пользователей.

Описанные психологические ограничения приводят к появлению характерных значений у величины k . Их можно описать через моменты распределения

$$\langle k^\alpha \rangle = \frac{1}{N} \sum_{i=1}^N k_i^\alpha \cong \int k^\alpha \cdot u(k) dk.$$

Отношение последовательных (с показателями, различающимися на единицу) моментов имеет ту же размерность, что и исходная величина. Врезка к рис. 1 дает представление об изменении этого отношения. В частности, отношение первого момента к нулевому определяет среднюю степень вершины $\langle k^{\text{in}} \rangle = \langle k^{\text{out}} \rangle \approx 18,1$ для направленных и $\langle k \rangle \approx 10,1$ для взаимных связей. При возрастании показателя отношение стремится к максимальному зарегистрированному значению, равному 63,5 тыс. для входящих, 17,7 тыс. для исходящих и 6,2 тыс. для взаимных связей. Легко видеть, что и при промежуточных значениях показателя первых намного больше, чем вторых, а вторых – чем третьих.

2. Вершинная корреляция, ассортативные свойства

Исследование социальных сетей Марком Ньюманом в 2002 году показало, что в некоторых ситуациях вершины с высокой степенью имеют склонность связываться с другими вершинами, также имеющими высокие степени. В этом случае сеть показывает так называемое *ассортативное* смешивание. Если, напротив, вершины, имеющие высокие степени, присоединяются к вершинам с низкими степенями, то такое смешивание называется *дисассортативным* [2].

Исследования показали, что социальные сети относятся к ассортативным сетям, в то время как технологические сети (Интернет, WWW) – к дисассортативным. Причины появления таких свойств еще полностью не изучены.

Для того чтобы определить, какими свойствами обладает сеть Живого журнала, необходимо посчитать корреляцию между вершинами – насколько две вершины схожи между собой. Для определения наличия корреляций в сети была использована *средняя степень ближайших соседей вершины* K_{nn} (индекс nn – от английского “nearest neighbors” – означает осреднение по ближайшим соседям) [3]. Ее значение для вершины i для связей разного типа вычисляется по формулам

$$K_{nn,i}^{\text{out}} = \frac{1}{k_i^{\text{out}}} \sum_{j=1}^N v_{ij} k_j^{\text{out}}$$
$$K_{nn,i}^{\text{in}} = \frac{1}{k_i^{\text{in}}} \sum_{j=1}^N v_{ji} k_j^{\text{in}} \quad .$$
$$K_{nn,i} = \frac{1}{k_i} \sum_{j=1}^N w_{ij} k_j$$

Интерес представляет изменение величины K_{nn} при возрастании степени рассматриваемой вершины k . Если средняя степень ближайших соседей возрастает, то вершины с высокой степенью (хабы) связаны между собой, и, наоборот, вершины с низкой степенью связываются с вершинами низкой степени – т.н. *ассортативное смешивание* [2].

С другой стороны, если K_{nn} убывает с ростом k , то большинство связей хабов ведут к более или менее изолированным вершинам, т.е. вершины с высокой степенью связываются с вершинами с низкой степенью. И в этом случае сеть демонстрирует *дисассортативное смешивание*.

Если K_{nn} не изменяется от степени вершин, то такая сеть называется *некоррелированной*.

Для безмасштабной сети следует ожидать зависимости вида

$$K_{nn} \sim k^r,$$

т.е. характер смешивания в сети определяется знаком показателя r .

Рис. 2 демонстрирует, что в случае Живого журнала для пользователей, имеющих от нескольких десятков до примерно тысячи связей, смешивание является ассортативным, т.е. общительные блогеры предпочитают дружить с общительными. Однако вне указанного диапазона k сеть является некоррелированной или даже слабо-дисассортативной. Это вполне естественно, поскольку новички очень часто устанавливают первые связи с завсегдатаями.

Следует отметить, что степень вершины является не вполне самостоятельной характеристикой. Некоторые ее значения являются психологически обусловленными. Например, престижно быть «тысячником», т.е. быть другом

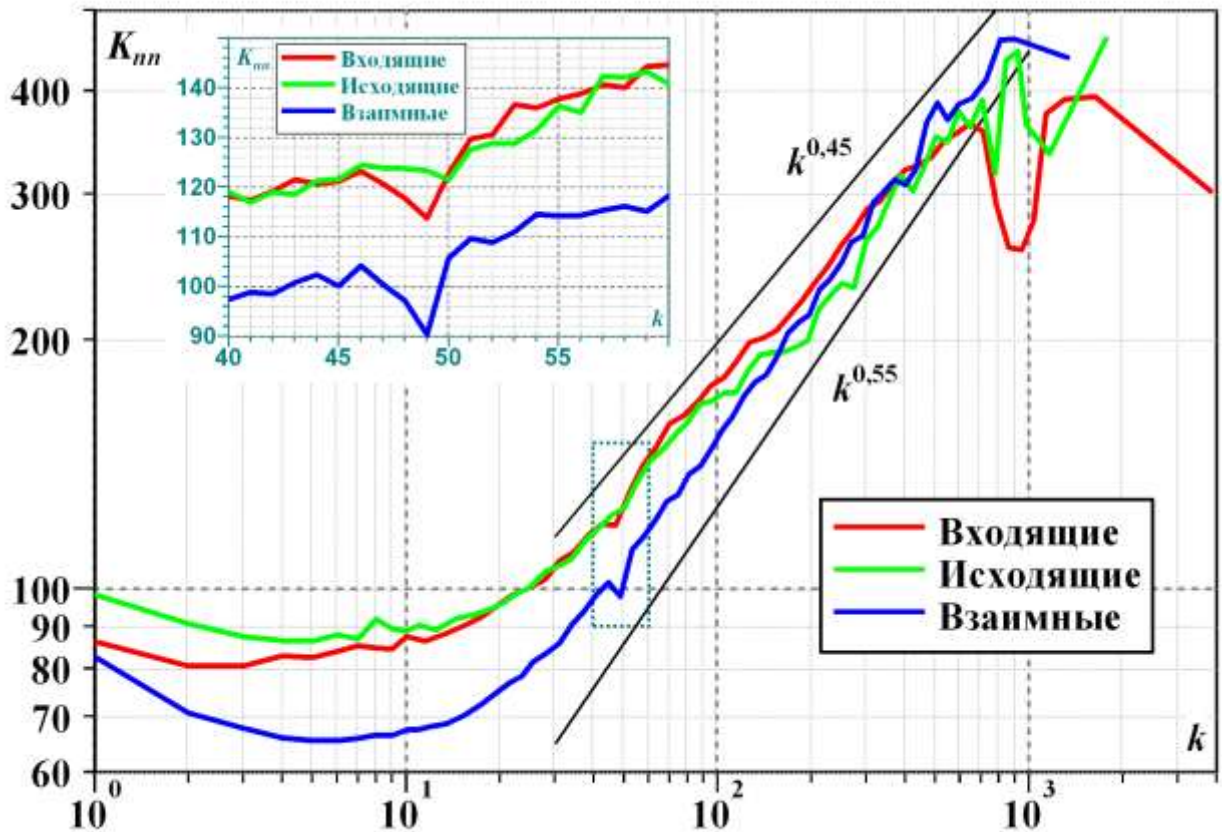


Рис. 2. Средняя степень ближайших соседей вершины

Область промежуточной асимптотики, в которой зависимости для всех типов связей характеризуются показателем $r \approx 0,45 \div 0,55$, имеет ширину примерно полтора порядка. Вне этой области r близко к нулю или даже отрицательно.

Обращают на себя внимание провалы графиков в районе значений $k \approx 50$ и 1000 . Область первого из них, обведенная пунктирной рамкой, увеличена на врезке.

тысячи других пользователей. Поэтому неудивительно, что блогеры, подбирающиеся к заветному рубежу, стремятся к тому, чтобы их зафрендил как можно больше народу. Из-за этого график $K_m^{in}(k^{in})$ имеет отчетливый провал в районе $k^{in} \approx 1000$. Здесь ассортативность смешивания перестает действовать – становится важно уже не то, кто с тобой дружит, а всего лишь то, сколько их.

Похожий, хотя и значительно менее выраженный, провал имеется для $K_m(k)$ и $K_m^{in}(k^{in})$ при степени вершины, чуть меньшей 50 (см. врезку на рис. 2). По всей видимости, этот рубеж также психологически важен. Однако у новичка еще мало шансов, что его зафрендит кто-то совсем незнакомый, поэтому эта полусотня состоит из взаимных друзей – тех, кого сначала зафрендил он сам.

3. Структура вторичных связей

Еще одной важной характеристикой вторичных связей является *коэффициент кластеризации* (the clustering coefficient). Эта характеристика, введенная Дунканом Уоттсом и Стивеном Строгатцом в 1998 г., определяется как средняя доля пар соседей вершины, которые являются также соседями друг друга. При-

менительно к ЖЖ величина коэффициента кластеризации говорит о том, как тесно связан круг виртуальных друзей человека.

Для вершины i коэффициент кластеризации

$$C_i = \frac{G_i}{k_i(k_i - 1)/2},$$

где G_i – количество связей между ее соседями

$$G_i^{\text{in}} = \sum_{j,m=1}^N v_{ji} v_{mi} \cdot \frac{v_{jm} + v_{mj}}{2}$$

$$G_i^{\text{out}} = \sum_{j,m=1}^N v_{ij} v_{im} \cdot \frac{v_{jm} + v_{mj}}{2}.$$

$$G_i = \sum_{j,m=1}^N w_{ij} w_{im} \cdot w_{jm}$$

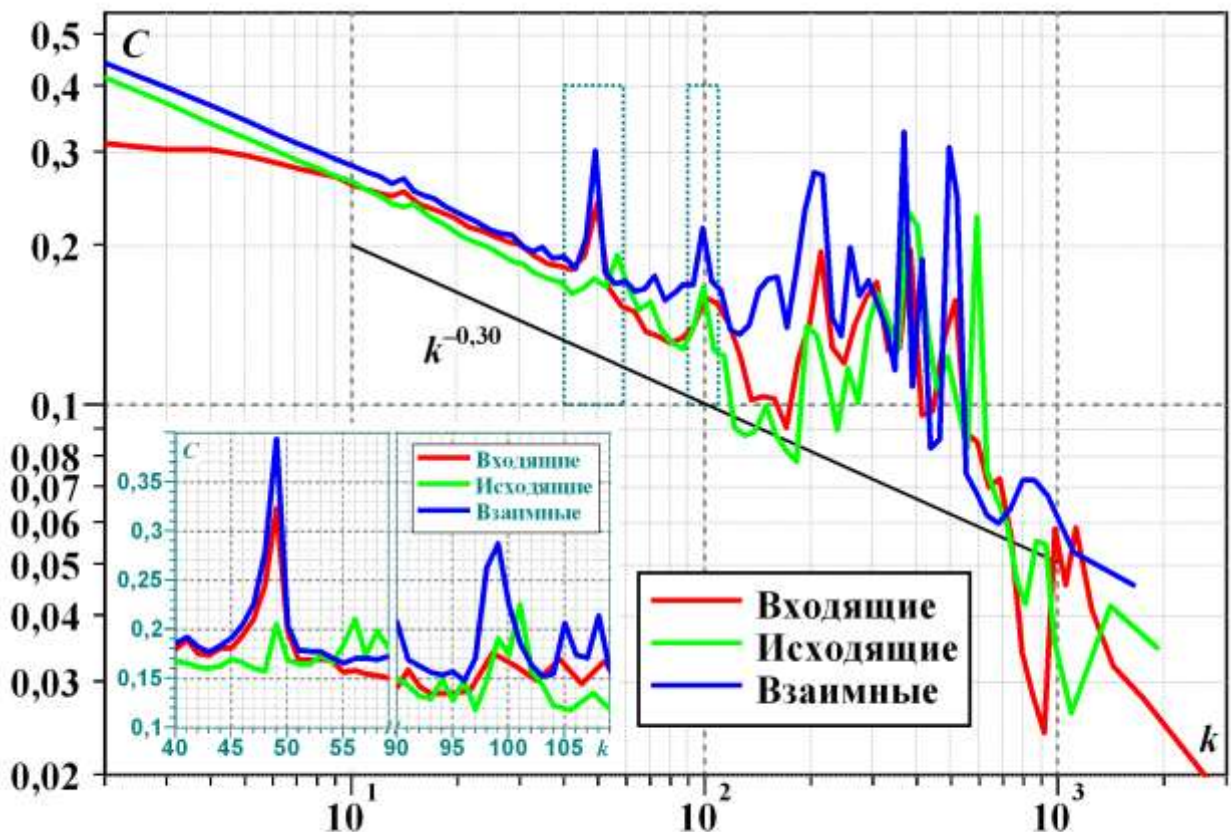


Рис. 3. Коэффициент кластеризации вершины в зависимости от ее степени

При в целом степенном характере зависимости на ее график накладываются резкие выбросы. В области сравнительно небольших k они приходятся на «красивые» значения около 50, 100, 200, а далее приобретают характер шума.

На врезке увеличены области двух выбросов, обведенных пунктирной рамкой. Видно, что на самом деле пики расположены непосредственно перед круглыми значениями степени.

Средние значения коэффициента кластеризации для Живого журнала составляют 25,8% для входящих, 27,6% для исходящих и 31,3% для взаимных связей, что немало. Наибольшее значение коэффициента именно для взаимных связей обусловлено тем, что такие связи в социальной сети по смыслу ближе всего к обычной дружбе, при которой друзья одного человека имеют прекрасную возможность познакомиться и друг с другом.

Следует отметить, что одинаковый учет всех вершин при осреднении не вполне корректен и искажает картину. При таком учете необоснованно высокое значение придается многочисленным вершинам с малой степенью, которые, как мы видели выше, еще плохо подчиняются статистическим свойствам безмасштабной сети. Более разумным представляется проводить осреднение, учитывая каждую вершину с весом, равным ее степени. При этом средние значения коэффициента кластеризации уменьшаются до 15,3% для входящих, 15,4% для исходящих и 20,8 % для взаимных связей.

Рис. 3 демонстрирует общую тенденцию уменьшения кластеризации соседей вершины по мере увеличения ее степени. Однако непосредственное использование зависимости $C(k)$ для анализа неудобно из-за выбросов, обуслов-

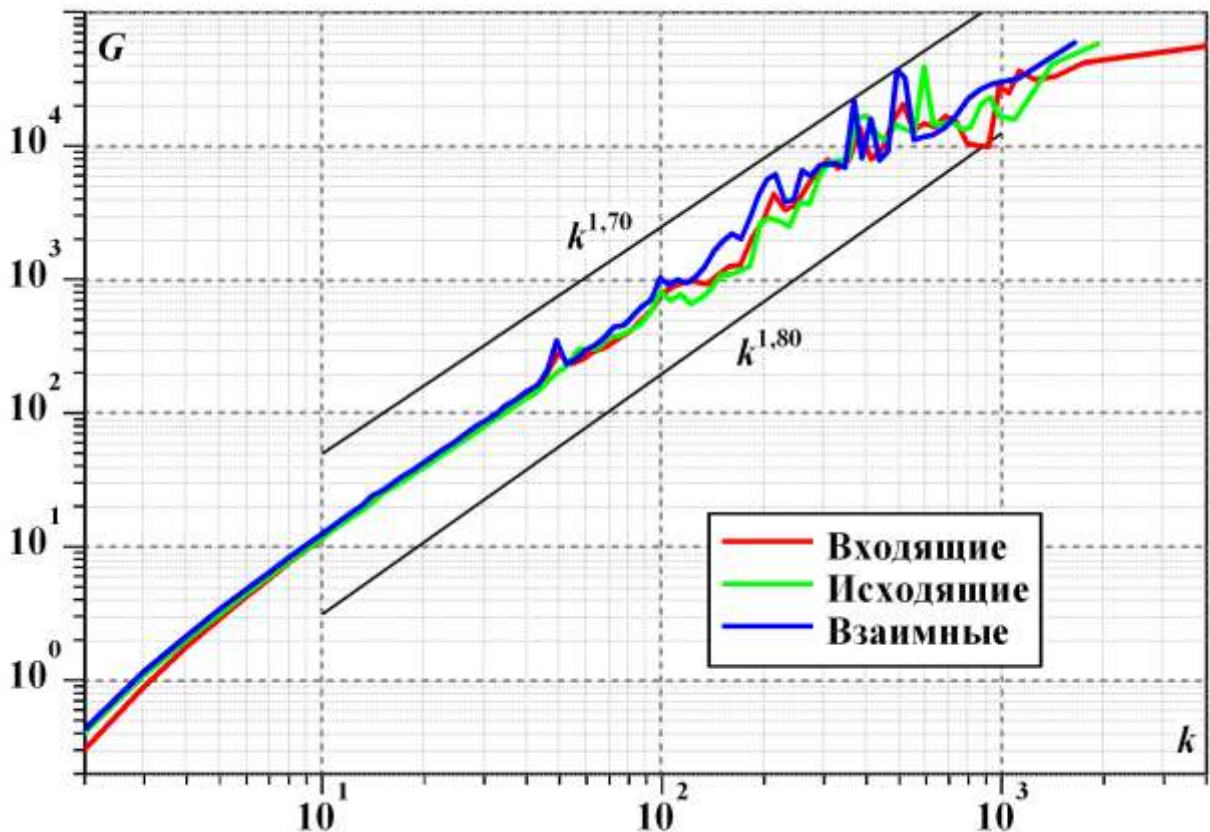


Рис. 4. Количество связей между соседями вершины в зависимости от ее степени

В области значений степени шириной два порядка зависимость характеризуется показателем $\gamma = 1,7 \div 1,8$.

Даже на зависимости $G(k)$ видны пики в районе значений $k \approx 50$, 100 и 200. Но они намного менее выражены, чем на рис. 3, хотя графики на обоих рисунках построены с одними и теми же параметрами сглаживания.

ленных психологией пользователей, и сильного влияния флуктуаций.

Более удобным является рассмотрение величины G_i , для которой эти эффекты менее заметны. Ее график, показанный на рис. 4, выглядит значительно опрятнее. Он подтверждает масштабно-инвариантную зависимость среднего числа связей между соседями вершины от ее степени

$$G(k) \sim k^\gamma,$$

которую следует ожидать для безмасштабной сети. Рис. 4 подтверждает такой характер зависимости. Ее показатель $\gamma < 2$ на величину, превышающую погрешность его определения. Таким образом, средний коэффициент кластеризации соседей вершины убывает с ростом ее степени как

$$C(k) \sim k^{\gamma-2},$$

что означает разрежение круга друзей по мере увеличения их числа.

Обратим внимание на близость графиков для разных типов связей на рис. 3 и особенно на рис. 4. Структура вторичных связей в меньшей степени зависит от их типа, нежели топология сети или вершинная корреляция.

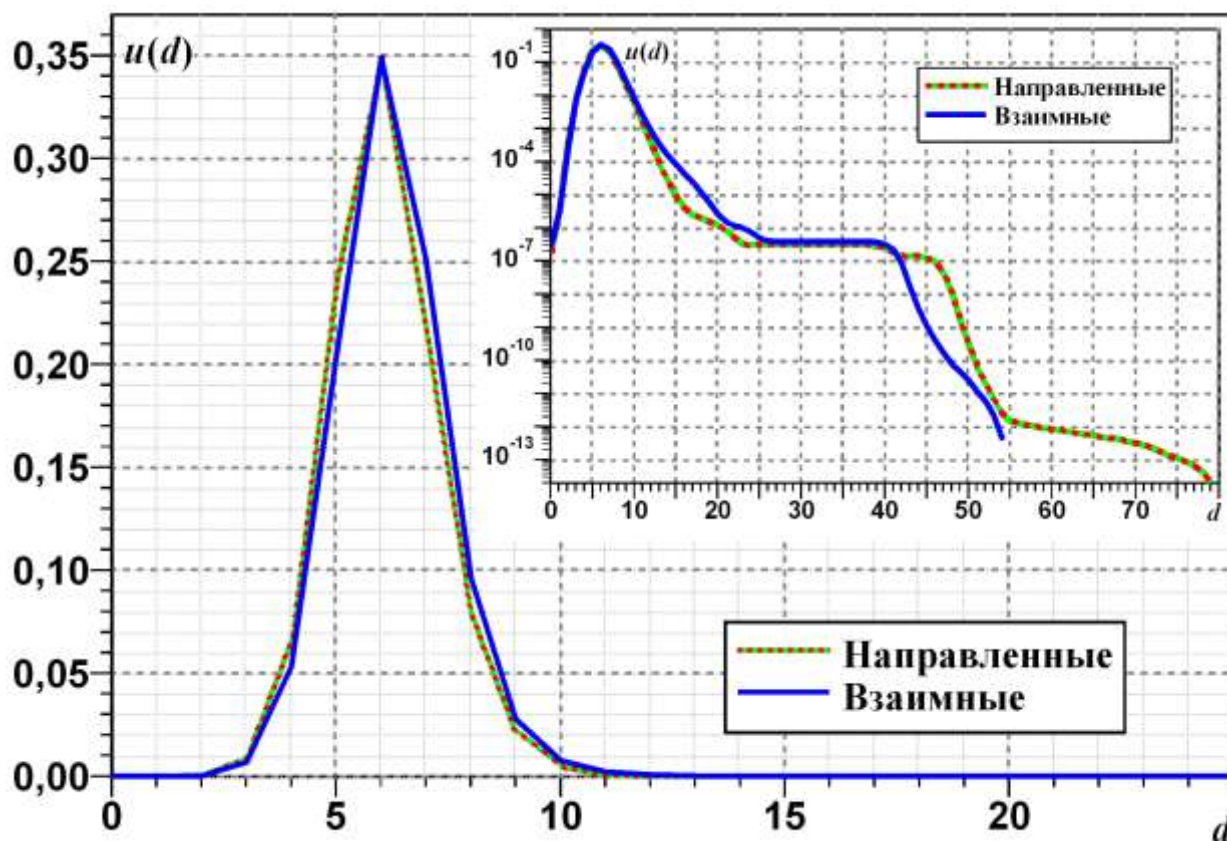


Рис. 5. Распределение пар вершин по кратчайшему расстоянию между ними

Плотность отнормирована на вероятность того, что между вершинами существует путь, составляющую 70,1% для направленных и 42,5% для взаимных связей.

На врезке те же графики построены в полулогарифмическом масштабе.

4. Кратчайшие пути и малые миры

Кратчайшие пути играют важную роль в транспортировке и коммуникациях внутри сети, а также в определении свойств структуры графа. Наименьшее число связей, необходимое для перехода от узла i к узлу j , называется *расстоянием* d_{ij} между двумя вершинами i и j (длины связей между смежными вершинами считаются равными единице). Максимальное значение d_{ij} называется диаметром графа. Если число вершин в графе велико, а расстояния между ними – нет, то говорят об эффекте *малого мира*.

На рис. 5 показано распределение $u(d)$ расстояний между случайно выбранными вершинами для Живого журнала. График имеет острый пик в районе значения $d = 6$. Свыше 95% связанных пользователей отстоят друг от друга на $4 \div 8$ звеньев сети, что подтверждает концепцию шести степеней разделения С. Милгрэма.

Следует заметить, что диаметр графа оказывается значительно больше приведенного значения, составляя 79 для направленных и 54 для взаимных связей. Однако доля пар вершин, разнесенных на столь большое расстояние, ни-

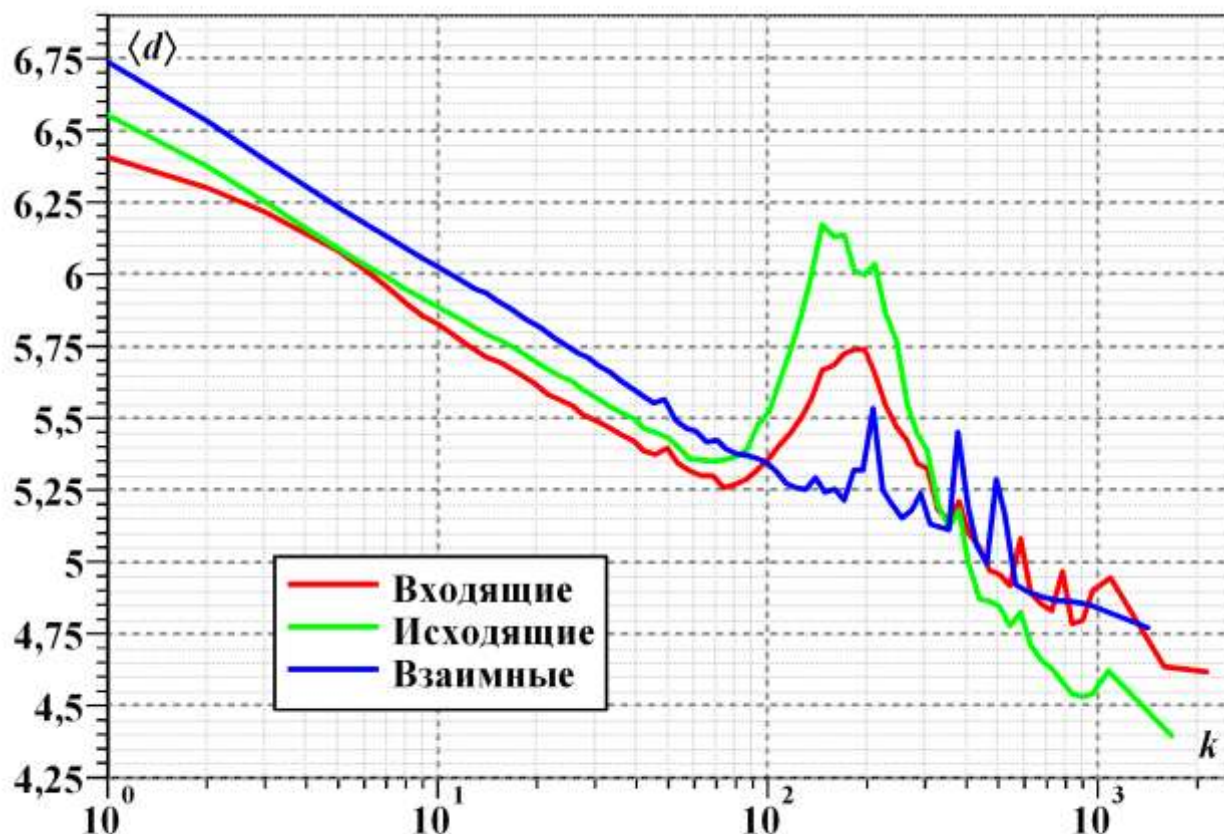


Рис. 6. Средняя удаленность вершины от сети в зависимости от степени

Увеличение степени вершины на полтора порядка приближает ее к сети на одно звено.

Природа горбов для направленных связей, возникающих при степени свыше сотни, неизвестна. Можно предположить, что этот тот порог, за которым пользователь становится интересен тем, кто пока находится на периферии сети, и сам начинает проявлять интерес к ним.

чтожна. Чтобы его зарегистрировать, необходимо оба конца пути выбрать на торчащих из общего облака связей длинных цепочках, которые являются весьма экзотическими объектами. Разумно вообще пренебречь встречаемостью пар редких настолько, что вероятность их выпадения становится соизмерима с вероятностью случайно выбрать одну и ту же вершину в качестве обоих концов пути. В этом случае можно считать $d_{\max} \approx 25$. Свыше 99,9995% пар вершин, между которыми имеется путь по графу, разделены расстоянием, не превышающем указанной величины.

Как можно видеть из рис. 6, средняя удаленность вершины от остального их множества логарифмически убывает с ростом ее степени, т.е. «притиснуться» к тесному миру чуть теснее оказывается очень дорогой задачей – для этого нужно существенно увеличить число связей. В безмасштабной сети всё находится очень близко, буквально на расстоянии вытянутой руки, но не ближе.

УСТОЙЧИВОСТЬ ЖИВОГО ЖУРНАЛА

Первые работы по исследованию устойчивости сетей были проделаны Альбертом-Ласло Барабаши в его статье [5]. Барабаши со своей группой в январе 2000 г. провел ряд компьютерных экспериментов с целью исследования устойчивости Интернета к отказам маршрутизаторов. При помощи имеющейся карты Интернета они случайным образом удаляли узлы в сети. Постепенно увеличивая долю удаленных узлов, они определяли критический уровень воздействия, при котором Интернет распадается на несвязанные части. К их большому удивлению сеть отказывалась ломаться. Даже при удалении 80% всех узлов, оставшиеся 20% все еще формировали связанную группу. Это открытие показало, что Интернет, в отличие от многих других систем, показывает высокую степень устойчивости против случайных отказов отдельных узлов – маршрутизаторов.

Действительно, исследования Анн-Арбора из университета Мичигана показали, что в любой момент времени происходят сотни сбоев маршрутизаторов, но, несмотря на частые и неизбежные поломки, пользователи редко замечают существенные разрушения Интернета. При исследовании устойчивости масштабно-инвариантных сетей оказалось, что они показывают такие же свойства: можно случайным образом удалить существенную долю узлов сети без нарушения ее целостности [6].

Возникает вопрос: все ли масштабно-инвариантные сети одинаково устойчивы к случайным отказам? Ответ на этот вопрос дал С. Хэвлин, который решил эту задачу на основе теории перколяции. Вместе с Р. Коэном, К. Эрезом и Д. Бенон-Авраханом он вычислил критическую долю удаленных узлов в случайных и масштабно-инвариантных сетях. С одной стороны, полученные ими результаты объясняли тот факт, что случайные сети полностью распадаются после удаления некоторой критической доли узлов. С другой стороны, они определили, что для масштабно-инвариантных сетей критический порог исче-

зает. Удивительно, что большинство сетей от Интернета до клетки, являются масштабно-инвариантными.

Группа исследователей под руководством А. Барабаши изучала также устойчивость сети к направленным атакам на вершины. В этих экспериментах вершины (хабы) удалялись из сети одна за другой. Удаление первого хаба не нарушало целостности системы, т.к. остальная их часть была все еще способна скрепить сеть, но после удаления нескольких хабов разрушительный эффект становился очевидным. Продолжая удалять все больше и больше хабов, можно наблюдать развал сети [5]. Удаление небольшого числа хабов превращает сеть во множество крошечных изолированных частей.

Масштабно-инвариантные сети скрывают в своей структуре ахиллесову пяту – высокая устойчивость к случайным отказам сочетается с чрезвычайной уязвимостью к направленным атакам. Для того чтобы разрушить масштабно-инвариантные сети, достаточно удалить примерно 5–15% хабов. Таким образом, для остановки работы Интернет необходимо вывести из строя всего-навсего несколько сотен маршрутизаторов.

Кроме того, результаты исследования устойчивости ассортативных и дисассортативных сетей показывают, что по отношению к случайным отказам данные сети ведут себя практически одинаково, а по отношению к направленным атакам устойчивость дисассортативных сетей оказывается значительно меньше, чем ассортативных [7].

Несмотря на то, что основное содержание Живого журнала – личные записи, читаемые лишь друзьями, у него развилась важная побочная функция, связанная с публикацией широковещательных сообщений, обычно связанных с критическими ситуациями или общественно-значимыми темами. Такие сообщения, обычно размещаемые с пометками типа «прошу о перепосте» или «перепост приветствуется», копируются некоторыми увидевшими их пользователями, быстро распространяясь по сети. Поскольку многие из таких широковещательных сообщений в последнее время становятся элементом политической борьбы или самоорганизации гражданского общества, представляет интерес устойчивость ЖЖ как среды их распространения по отношению к направленным атакам на его узлы. Не менее важным является и устойчивость по отношению к случайным отказам, основной причиной которых может являться массовое нежелание пользователей воспроизводить те или иные сообщения.

Устойчивость ЖЖ как среды распространения по отношению к направленным атакам и случайным отказам исследовалась двумя способами: случайным удалением узлов из сети и направленным воздействием на сеть. В первом случае из сети удалялась доля вершин s , выбранных случайным образом. Во втором – удалялись вершины с максимальной степенью, которые обычно выступают в роли если и не непосредственных источников, то усилителей широковещательных сообщений.

Определялась еще одна характеристика сети – размер компоненты, позволяющей получить представление о глобальной топологии сети. При помощи данной характеристики можно определить, является ли граф связанным или

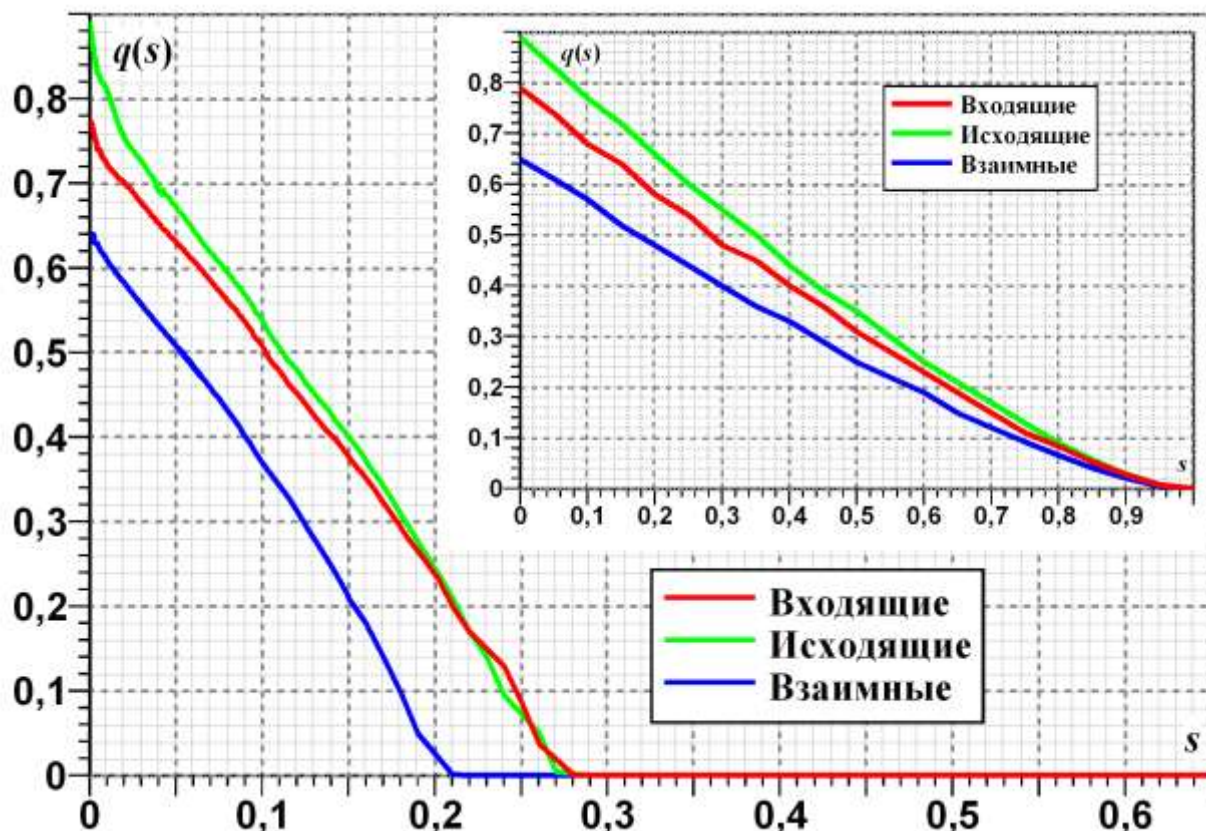


Рис. 7. Доля вершин в гигантской компоненте при удалении части вершин

Основной график – направленные атаки, врезка – случайные отказы.

В первом случае при удалении определенной доли крупнейших вершин гигантская компонента разрушается. Во втором случае гигантская компонента сохраняется до самого конца.

нет, и, если нет, то из скольких компонент он состоит. *Гигантской компонентой* называется такая компонента, которая имеет максимальный размер.

Так как сеть ЖЖ направленная, определение размера компоненты становится более сложной задачей. В данном случае сеть разделяется на компоненты сети для исходящих и входящих связей.

На рис. 7 представлены графики зависимостей $q(s)$ доли вершин, принадлежавших гигантской компоненте для разных типов связей, от доли удаленных вершин s . Эти графики позволяют судить об устойчивости сети по отношению к направленным атакам на узлы и к случайным отказам отдельных узлов.

Исследование устойчивости сети к случайным отказам некоторых пользователей ЖЖ для взаимных, входящих и исходящих связей показало, что она практически не меняется и сеть разбивается на мелкие изолированные узлы при отказе примерно 95% узлов, т.е. ее связанность сохраняется практически до самого конца.

Противоположный эффект имеет место в случае направленных атак. Сети исходящих и входящих связей распадаются на несвязные части при удалении около 27% хабов. Для сети взаимных связей пороговое значение еще меньше и составляет лишь 21%.

ВЫВОДЫ

Исследование свойств Живого журнала продемонстрировало масштабно-инвариантный характер его сети, выражающийся в степенном виде распределения вершин по числу связей, а также зависимостей от степени вершины средней степени ее соседей и их кластеризации.

Плотность распределения кратчайших путей между случайными парами вершин характеризуется очень резким пиком, расположенным около шестерки, что подтверждает концепцию шести степеней разделения С. Милгрэма и позволяет отнести ЖЖ к малым мирам.

Продемонстрирован пороговый характер уязвимости ЖЖ как среды распространения широкоэвентальных сообщений по отношению к случайным отказам и направленным атакам. В первом случае устойчивость оказалась очень высокой (с порогом 0,95), во втором – сравнительно низкой (с порогом $0,2 \div 0,3$).

Литература

1. Milgram S. The small world problem // *Psychology Today*. 1967. № 2. P. 60–67.
2. Newman M. E. J. Assortative mixing in networks/ Department of Physics, University of Michigan, Ann Arbor, MI 48109–1120 and Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501.
3. Pator-Satorras R., Vazquez A. and Vespignani A., *Phys. Rev. Lett.* 87 (2001), p.258701.
4. Watts D.J. and Strogatz S.H., *Nature* 393 (1998), p.440.
5. Albert R., Jeong H., and Barabási A.-L. Error and attack tolerance in complex networks, *Nature* 406, 378 (2000).
6. Barabási A-L. *Linked: How everything is connected to everything else and what it means for business, science and everyday life*, USA (2002).
7. Гаджиев Б.Р., Прогулова Т.Б., Щетинина Д.П. Статическая устойчивость ассортативных и дисассортативных сетей/ *Математика, компьютер, образование. Сборник научных трудов. Выпуск 14. Том 2/ Под ред. Г.Ю. Резниченко.* — М. – Ижевск: НИЦ «Регулярная и хаотическая динамика». 2007, с 22-29.

Содержание

Введение.....	3
Структура Живого журнала	3
1. Топология сети	4
2. Вершинная корреляция, ассортативные свойства	6
3. Структура вторичных связей	8
4. Кратчайшие пути и малые миры	12
Устойчивость Живого журнала	13
Выводы	16