



ИПМ им.М.В.Келдыша РАН • Электронная библиотека

Препринты ИПМ • Препринт № 107 за 2014 г.



Митин Н.А., Щетинина Д.П.

**Исследование свойств сети
сотрудничества киноактеров**

Рекомендуемая форма библиографической ссылки: Митин Н.А., Щетинина Д.П.
Исследование свойств сети сотрудничества киноактеров // Препринты ИПМ им. М.В.Келдыша.
2014. № 107. 20 с. URL: <http://library.keldysh.ru/preprint.asp?id=2014-107>

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

Н.А.Митин, Д.П.Щетинина

**Исследование свойств сети
сотрудничества киноактеров**

Москва — 2014

Н.А.Митин, Д.П.Щетинина

Исследование свойств сети сотрудничества киноактеров

В данной работе исследуется структура связей в социальной сети сотрудничества киноактеров России за период с 1908 по 2013 год. Проанализированы ее основные характеристики: распределение вершин по степени, корреляции при образовании связей, кластеризации вторичных связей. Выявлен масштабно-инвариантный характер некоторых количественных характеристик. Исследование малых миров при помощи распределения кратчайших путей вершины показало, что актерский мир в два раза уже, чем шестистепенная концепция разделения С. Милгрэма. В конце рассматривается динамика сети сотрудничества киноактеров.

Ключевые слова: социальные сети, малые миры, масштабная инвариантность, степенные распределения, вершинная корреляция, ассортативность, локальная кластеризация и избыточность.

N.A.Mitin, D.P.Shchetinina

The study of movie actor collaboration properties

We investigate the connection structure in social movie actor network of the period from 1908 to 2013. We search the main quantitative properties of its network such as the degree distribution, correlations and the clustering coefficient. Some of these dependences are found to be scale invariant. Research of small worlds by means of distribution of the shortest paths showed that the actor's world is twice less, than Milgram's concept of six degrees of separation. In addition we search the dynamics of social movie actor collaboration network.

Key words: small-worlds, social networks, scale invariance, power laws, vertex correlation, assortativity, the clustering of secondary connections.

Работа выполнена при поддержке РФФИ (проекты 12-06-00402 и 13-01-00617).

Оглавление

Введение	3
Структура сети сотрудничества киноактеров РФ	4
Динамика сети сотрудничества киноактеров	15
Выводы	19
Библиографический список.....	20

ВВЕДЕНИЕ

В последнее время возрос интерес к исследованию социальных сетей. Многие люди, когда говорят о социальных сетях, часто подразумевают онлайн сети, типа Вконтакте, Facebook, Мой мир, Живой журнал [1] и др. Однако само понятие социальные сети ведет в прошлое и означает значительно больше, чем просто современные компьютерные воплощения сети. Достоверные исследования в этой области обычно относят к работам психиатра Якоба Морено (Jacob Moreno) – иммигранта из Румынии, переехавшего в США и с 1930 года интересовавшегося динамикой социальных взаимодействий внутри групп людей. В марте 1933 г. на конференции по медицине он представил результаты своих исследований, которые в дальнейшем стали первым достоверным результатами в изучении социальных сетей. Они были опубликованы в газете «Нью-Йорк Таймс». Год спустя он опубликовал книгу «Кому удастся выжить?», которая, хотя и не соответствует современным стандартам исследования социальных сетей, но содержит первые ростки раздела социометрии, который в дальнейшем стал называться анализом социальных сетей.

В данной работе под социальной сетью понимается множество людей, а иногда и групп людей, которые связаны между собой некоторыми формами социальных отношений, такими, например, как дружба, общие интересы, обмен новостями, совместное соавторство в публикациях и др. В настоящее время существует огромное разнообразие таких видов сетей. Это и сети соавторства научных работ, сети цитирования, сети киноактеров Голливуда, онлайн социальные сети: Живой журнал, Twitter, Facebook и многие другие.

Огромный интерес для исследования представляет отдельный вид социальных сетей – социальные сети сотрудничества. К нему можно отнести сети сотрудничества киноактеров, ученых и другие. Пионером в области исследования социальной сети сотрудничества на примере киноактеров Голливуда стал А.Л. Барабаш, который в 2000 г. сконструировал сеть на основе базы данных кинофильмов (<http://www.imdb.com>), содержавшей все кинофильмы с составом исполнителей с 1980 г. [2]. В этой сети вершинами являются актеры, и две вершины имеют общую связь, если соответствующие актеры играли в одном кинофильме. Эта непрерывно расширяющаяся сеть на 1998 г. содержала 225226 актеров. Распределение вершин по числу связей подчиняется степенному закону, который при больших k имеет вид $u(k) \sim k^{-2.3}$.

В 2009 г. профессор Б.Р. Гаджиев с группой аспирантов Международного университета природы, общества и человека «Дубна» попытались построить аналог сети киноактеров Голливуда на основе художественных кинофильмов Российского кинематографа [3]. В исследование были включены данные двух основных киностудий России за период с 1918 по 1992 гг.: Ленфильм (www.lenfilm.ru) и Мосфильм (www.mosfilm.ru), а также отдельно рассматривалась Центральная киностудия детских и юношеских фильмов им. М. Горького. Следует отметить, что исследование проводилось только для основных актеров,

т.е. актеров первого и второго плана. Данная сеть содержала 2910 фильмов с 6997 актерами для объединенной сети Мос- и Ленфильма и 3078 актеров для киностудии им. М.Горького. Полученное распределение, как и для сети киноактеров Голливуда, подчинялось степенному закону $u(k) \sim k^{-2,14}$ при $k > 6$.

В данной работе мы рассматриваем сеть сотрудничества киноактеров российского кинематографа с первого сохранившегося фильма (1908 г.) и по май 2013 года. Отдельно исследуются сеть, состоящая из художественных фильмов, а также из сериалов. Предметом исследования является изучение статистических свойств данных сетей. Проводится сравнительный анализ статистических свойств рассмотренных сетей с результатами, полученными в работе [3]. Изучается влияние эпизодических актеров на статистические свойства сети. Также исследуется эффект малого мира. В конце рассматривается динамика сети киноактеров РФ.

СТРУКТУРА СЕТИ СОТРУДНИЧЕСТВА КИНОАКТЕРОВ РФ

Данные для исследования сети сотрудничества киноактеров были собраны на сайте www.kino-teatr.ru. Рассматривается период времени с 1908 г. по май 2013 г. В этот период времени история кинематографа претерпела ряд изменений в своем развитии. Она включает в себя три ключевых периода: кинофильмы Российской Империи (1908-1917 гг.), советское кино (1918-1991 гг.) и российские фильмы (1992 г. – май 2013 г.). Всего в базу вошло 16690 кинофильмов. К ним относятся: короткометражные фильмы (1606 шт.), документальные фильмы (87 шт.), художественные фильмы (11525 шт.), телеспектакли (1315 шт.) и сериалы (2157 шт.).

Из всех данных наибольший интерес представляет рассмотрение отдельных выборок кинофильмов по художественным фильмам, сериалам и полной базе фильмов основных актеров, а также статистика, включающая эпизодических актеров, которые составляют 42 % по художественным фильмам, около 50 % для сериалов и 43 % для полной сети.

Данные, представленные на сайте, организованы в виде фильмотеки с актерами, принимающими участие в конкретном фильме. Такой вид отношений удобно представлять в виде графа киноактеров и фильмов, где ребро соединяет соответствующего актера и фильм, в котором он снимался. Такой граф называется *двудольным*, или *биграфом*, т.е. множество его вершин можно разбить на две части: $U \cup V = W$, $|U| > 0$, $|V| > 0$, так, что ни одна вершина в U не соединена с вершинами в U и ни одна вершина в V не соединена с вершинами в V . В нашем случае к первой части множества будут относиться актеры российского кинематографа – вершины $V_i, i=1,2,\dots,N$, где N – общее число актеров в выборке, к другой части – фильмы, в которых снимались соответствующие киноактеры – вершины $U_i, i=1,2,\dots,s$, где s – общее число фильмов. В данном графе все связи между различными типами вершин рассматриваются только как нена-

правленные. Связь (ребро) представляет актера, принимающего участие в фильме.

Для того чтобы определить, какое количество актеров принимает участие в фильме U_i , введем характеристику $u(h)$. На рис. 1 изображено распределение фильмов по числу задействованных актеров. Данное распределение описывается степенным образом $u(h) \sim h^{-\gamma}$ с ростом увеличением числа задействованных

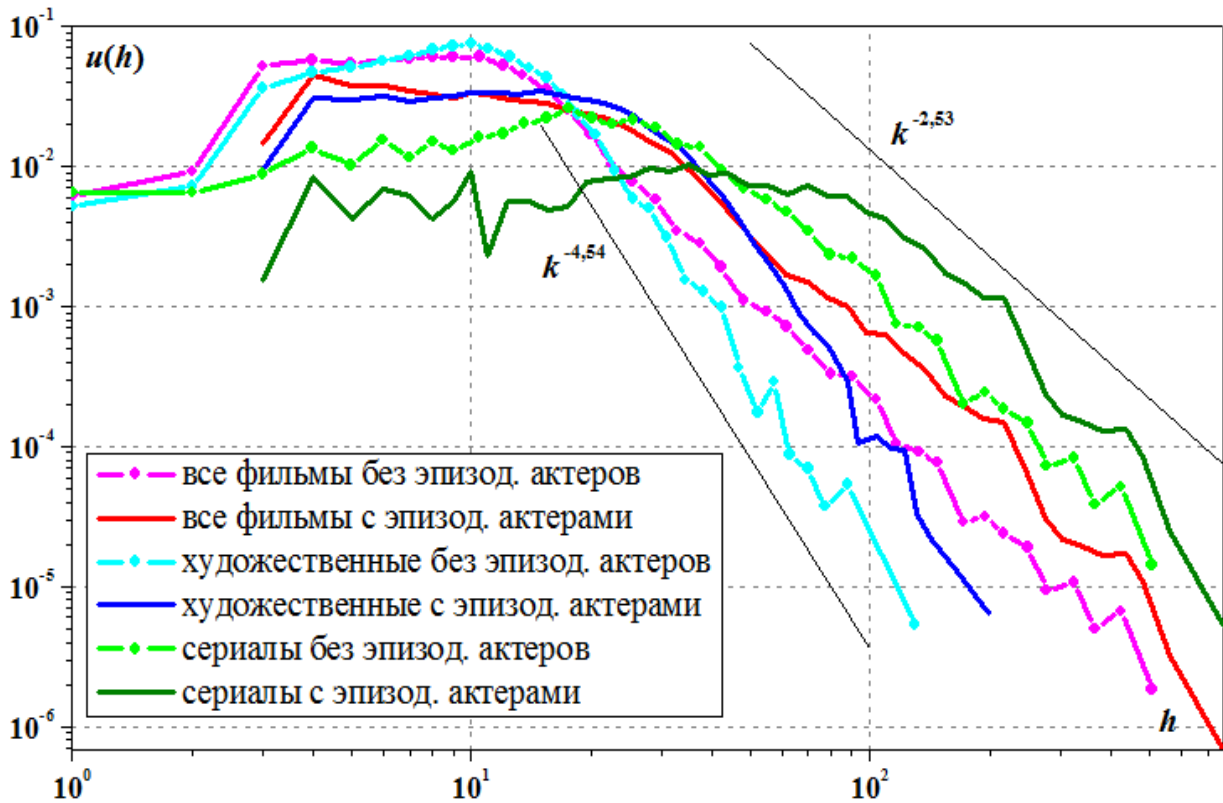


Рис. 1. Распределение фильмов по числу задействованных актеров

Плотность распределения имеет степенной хвост с показателем $\gamma = 2,56$ для полной сети, $\gamma = 4,54$ для художественных фильмов, $\gamma = 2,68$ для сериалов (эпизодические актеры включены) и, исключая эпизодических актеров, $\gamma = 2,73$ для полной сети, $\gamma = 4,31$ для художественных фильмов, $\gamma = 2,53$ для сериалов.

актеров в фильме.

Чтобы рассмотреть, во скольких фильмах актер i сыграл за свою жизнь, введем характеристику $u(m)$. На рис. 2 изображено распределение $u(m)$ актеров по числу сыгранных ролей по различным выборкам. Данное распределение $u(m) \sim m^{-\gamma}$ является степенным только в промежуточной асимптотике.

Однако, как показали ученые из Китая в работе [4], распределение $u(m)$ актеров по числу сыгранных ролей для некоторых сетей сотрудничества, в том числе и киноактеров Голливуда (с показателем $c = 0,45$), хорошо описываются растянутой экспонентой, которую впервые предложили французские ученые (J. Laherrere, D. Sornette) в 1998 г.

$$u(m) = \exp(-(m/m_0)^c), \text{ где } \ln u \sim m^c.$$

Очевидно, что если c близко к 1, распределение приближается к показательному распределению. Когда c близко к 0 – к степенному. Если c располагается между 0 и 1, то реализуется промежуточный вариант. Центральная часть эмпирического распределения, как может показаться, имеет линейную часть (масштабно-инвариантная область), однако в частях «головой» и «хвоста», как правило, есть искривления. Чем ближе значение c к 0, тем длиннее степенная область. В данном случае распределение также хорошо приближается растянутой экспонентой (см. рис. 3), где для полной сети $c = 0,4$, для сериалов $c = 0,55$ и для художественных фильмов $c = 0,33$ (в данных учтены эпизодические актеры). В случае исключения эпизодических актеров для полной сети получаем $c = 0,49$, для сериалов $c = 0,63$ и для художественных фильмов $c = 0,46$. Разница между показателями c составляет 0,09 для полной сети и сериалов, 0,13 – для художественных фильмов. Следует заметить, что показатель для сети из художественных фильмов без эпизодических актеров практически совпадает с показателем для сети сотрудничества Голливуда, разница с которым составляет

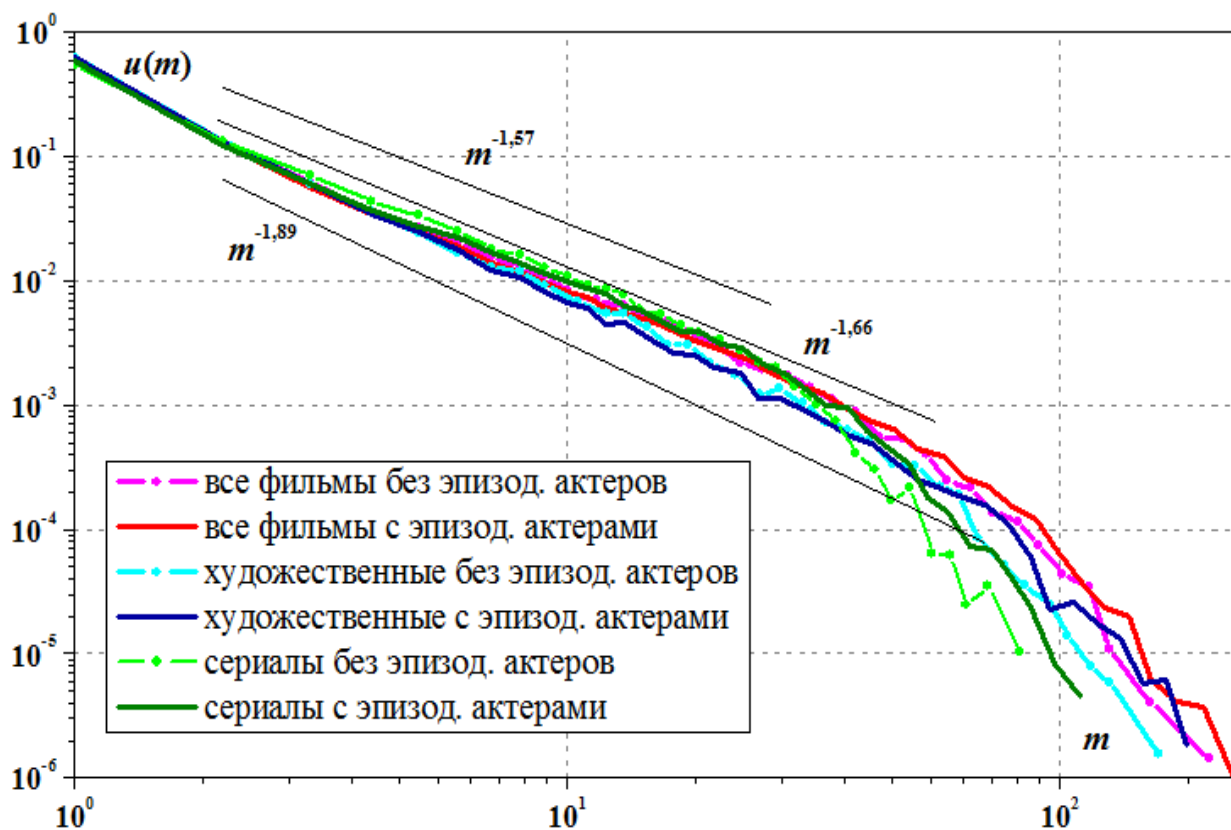


Рис. 2. Распределение актеров по числу сыгранных ролей

Плотность распределения описывается степенным законом в промежуточной асимптотике с $\gamma = 1,66$ для художественных фильмов, $\gamma = 1,89$ для полной сети и $\gamma = 1,57$ для сериалов, ширина которой составляет более одного порядка для сериалов и менее двух для художественных и полной базы фильмов.

Добавление эпизодических актеров к сети не влияет на изменение показателя степени.

0,01. При построении сети сотрудничества киноактеров рассматривают только одну часть множества V . Тогда связи с общей вершиной U_i двудольного графа

проецируются на связь между двумя узлами одного типа, и в этом случае связь l_{v_1} между V_1 и V_2 показывает сотрудничество актеров в фильме U_1 . Сеть удобно представлять в виде неориентированного графа, вершинами которого являются

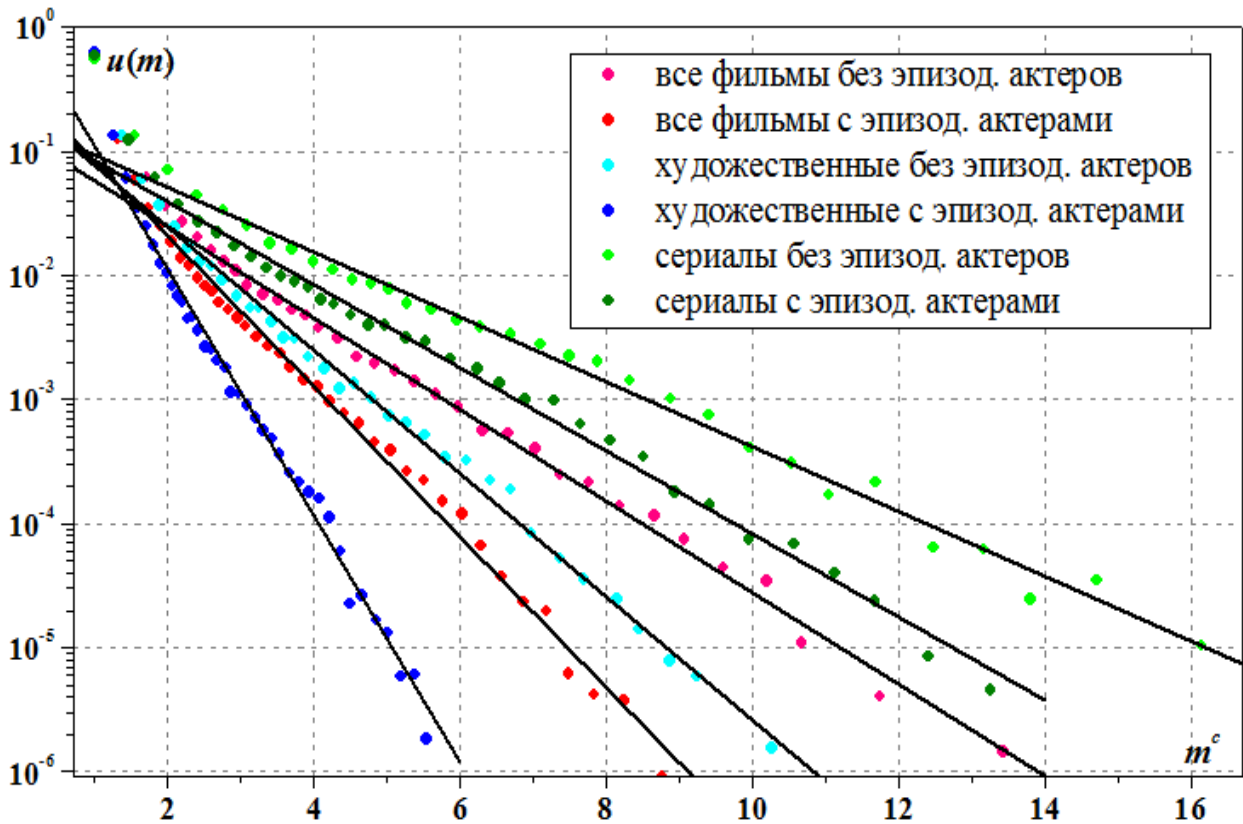


Рис. 3. Распределение актеров по числу сыгранных ролей в степени c

Плотность распределения описывается растянутой экспонентой на всем диапазоне значений, кроме первых двух точек.

актеры, и два актера связаны между собой ребром, если они сыграли в одном фильме хотя бы раз на протяжении своей карьеры. 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 i связана с вершиной j , и $v_{ij} = 0$ в противном случае. Все актеры фильма формируют полносвязный граф – *клик* (хотя, в действительности, зачастую бывает, что сцены актеров не пересекаются на протяжении всего фильма), и полная сеть состоит из множества полносвязных графов.

Полная сеть, рассматриваемая нами, с учетом эпизодических актеров состояла из 105 тыс. актеров с 35,5 млн. связей, средняя степень вершины $\langle k \rangle = 337$, без учета эпизодических актеров – из 60 тыс. с 11,44 млн. связей и $\langle k \rangle = 191$; сеть из художественных фильмов с учетом эпизодических актеров включает 71 тыс. актеров с 7,1 млн. связей и $\langle k \rangle = 101$, исключая эпизодических актеров – 41 тыс. с 2,2 млн. связей и $\langle k \rangle = 53$; сеть из сериалов с учетом эпизо-

дических актеров – 50 тыс. актеров с 28,6 млн. связей и $\langle k \rangle = 572$, без эпизодических актеров – 25 тыс. с 9,2 млн. связей и $\langle k \rangle = 366$.

1. Топология сети

Основной характеристикой топологии сети является распределение вершин по степени $u(k)$. Для масштабно-инвариантных сетей вероятность того, что вершина имеет *степень* k (число связей рассматриваемой вершины $k_i = \sum_{j=1}^N v_{ij}$), описывается степенным распределением вида

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

Данная формула, разумеется, применима лишь в определенном интервале значений – *промежуточной асимптотике*. При малых и больших значениях k происходит нарушение степенного закона. Рис. 4, на котором изображены распределения актеров по числу связей по различным выборкам с учетом эпизодических актеров, а также без них, позволяет в этом убедиться. Левая граница для полной базы фильмов лежит на уровне $k \sim 32$, для сети из художественных фильмов $k \sim 38$ и для сериалов на уровне $k \sim 122$ (эпизодические актеры учтены). Для выборок, исключая эпизодических актеров, левая граница для полной базы лежит на уровне $k \sim 13$, для художественных фильмов лежит на уровне $k \sim 16$, для сериалов $k \sim 35$. Для малых значений k это связано с тем, что актеров с малой степенью очень мало, и, скорее всего, они сыграли в одном небольшом фильме, например, в дипломной работе режиссеров, и больше нигде не снимались в своей жизни, т.е. «выпали» из профессии. Правая граница для сериалов лежит на уровне $k \sim 3000$, для художественных фильмов $k \sim 1700$, а для полной сети $k \sim 2200$ (эпизодические актеры включены). Для данных, исключая эпизодических актеров, для полной сети $k \sim 900$, для художественных фильмов $k \sim 600$ и для сериалов $k \sim 1000$. Это связано с тем что, никакая конечная система не может породить степенные законы распределения, применимые при любых значениях аргумента. Такие распределения характеризуются бесконечными моментами высоких порядков, что невозможно, т.к. $u(k)$ должно обратиться в нуль при достаточно большом k .

Небольшое значение показателя α (намного меньше 2) в распределении актеров по степени может объясняться следующими причинами. Продюсеры часто приглашают на съемки наиболее известных актеров, не рассматривая других кандидатов. Как правило, вовлечение в фильмы известных актеров увеличивает шансы на успех, так как зритель идет посмотреть на любимого актера. Таким образом, можно предположить, что *предпочтительный механизм* (богатый становится ещё богаче, или, другими словами, вероятность того, что новый актер окажется связанным ребром с уже присутствующим актером в сети, пропорционален его степени) должен доминировать, т.е. показатель будет равняться 2. Однако данные исследования охватывают столетний период времени,

поэтому ни у кого не может быть такой долгой артистической жизни. Таким образом, в данной сети невозможно существования центров (хабов), и, тем самым, не будет и предпочтительного механизма для целой сети. Но, если даже профессиональная жизнь актера и может продлиться в течение такого долгого времени, все равно не будет существовать глобального подавляющего *предпочтительного присоединения* из-за потребности в новых актерах, которые способны внести свежесть и новизну в снимаемые фильмы.

Добавление эпизодических актеров к сети играет важную роль в топологических свойствах сети. Происходит увеличение показателя степени на 0,11 для сети из художественных фильмов и полной базы фильмов и на 0,51 для сети из сериалов. Также происходит смещение левых границ на 19 единиц для полной сети, на 22 единицы для художественных фильмов и правых границ на 1100 единиц для художественных фильмов, на 1300 единиц для полной сети и 2000 для сериалов.

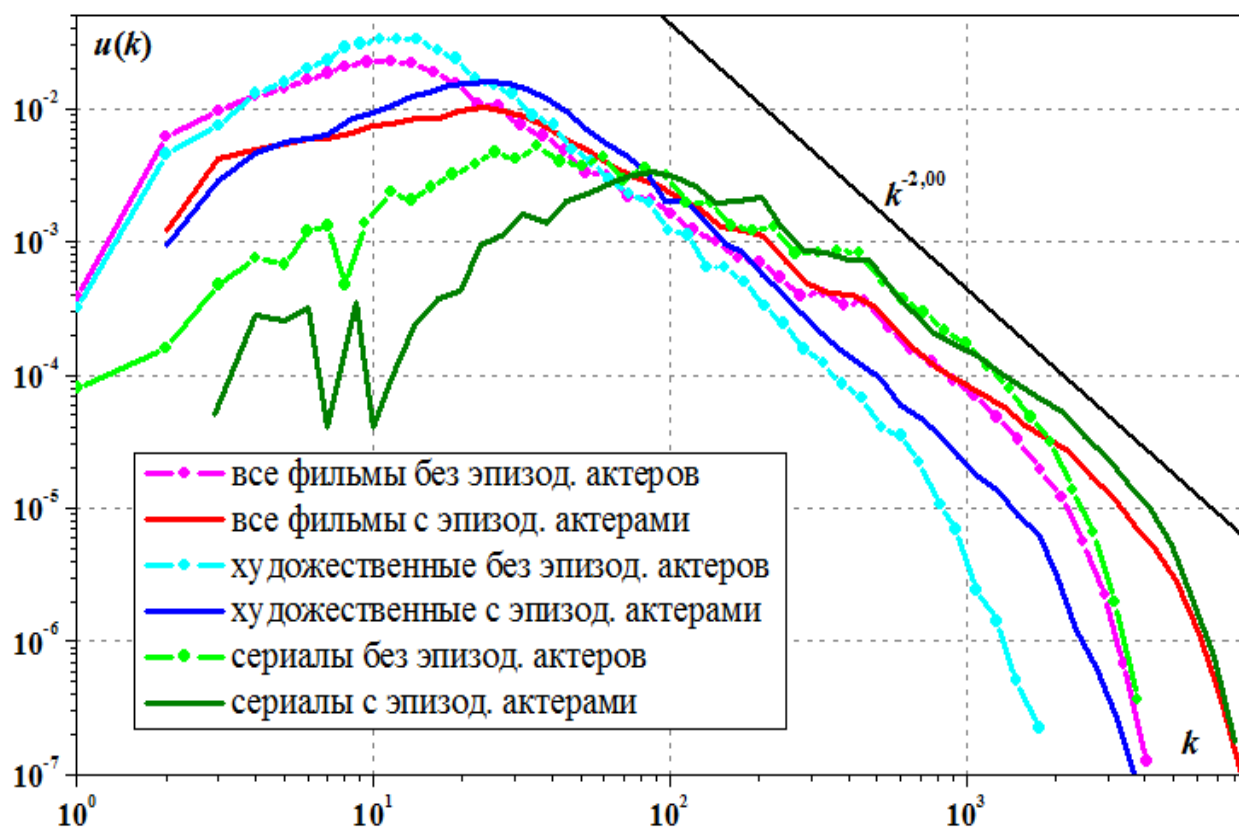


Рис. 4. Распределение актеров по числу связей

Плотность описывается степенным законом с показателем $\alpha \approx 1,00$ для сети из художественных фильмов, включая эпизодических актеров, $\alpha \approx 0,88$, исключая эпизодических актеров в промежуточной асимптотике, ширина которой составляет полтора порядка, для полной сети $\alpha \approx 0,38$, включая эпизодических актеров и $\alpha \approx 0,37$, исключая эпизодических актеров, ширина асимптотики составляет около двух порядков. Для сериалов с учетом эпизодических актеров $\alpha \approx 0,57$ и без учета эпизодических актеров $\alpha \approx 0,06$, ширина промежуточной асимптотике составляет около одного порядка.

Следует заметить, что показатель, полученный для художественных фильмов, исключая эпизодических актеров, оказывается меньше на 0,26 для

объединенной сети Мос- и Ленфильма по художественным фильмам основных актеров, исследованной в работе [3]. Можно предположить, что добавление одиночных актеров к сети влияет на увеличение показателя степени.

2. Свойства коррелированности. Ассортативность сети

Как правило, социальные сети *ассортативны*, т.е. вершины, имеющие близкие степени, связываются между собой [4].

Для того чтобы определить, какими свойствами обладает сеть киноактеров, необходимо посчитать корреляцию между вершинами. Для определения наличия корреляций в сети была использована *средняя степень ближайших соседей вершины* $K_{m,i}$ [5]. Ее значение для вершины i вычисляется по формуле

$$K_{m,i} = \frac{1}{k_i} \sum_{j=1}^N v_{ij} k_j .$$

Для исследования сети интерес представляет изменение величины $K_{m,i}$ 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* [4].

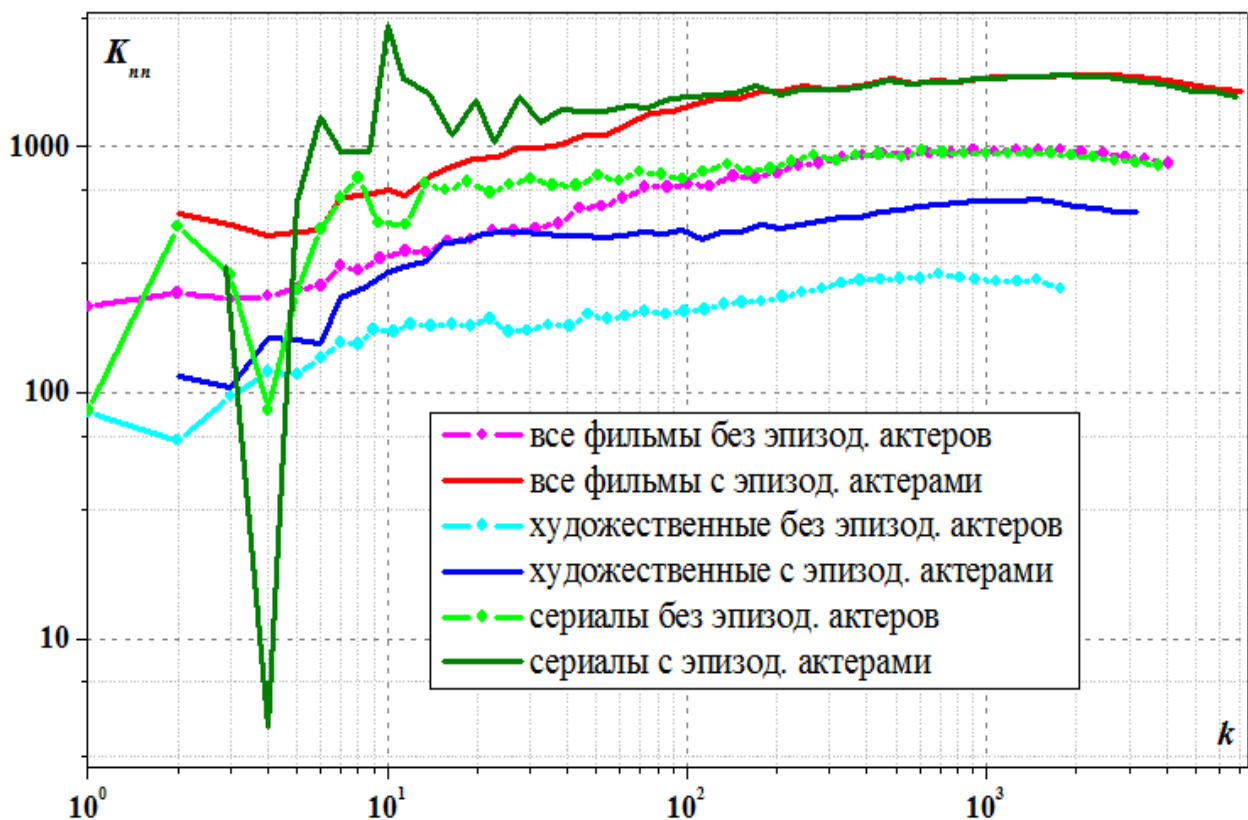


Рис. 5. Средняя степень ближайших соседей вершины

С другой стороны, если $K_{m,i}$ 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*. Технологические сети, такие как Интернет, WWW, относят к дисассортативным.

Если K_m не изменяется от степени вершин, то такая сеть называется *некоррелированной*.

Как было изучено на примере Живого журнала [1], для безмасштабной сети следует ожидать зависимости вида

$$K_m \sim k^r,$$

где характер смешивания в сети определяется знаком показателя r .

Но, как показано на рис.5, зависимость средней степени ближайших соседей вершины носит возрастающий характер, но не описывается степенным распределением. Возрастающая функция говорит о слабоассортативном смешивании сети для полной базы фильмов и для художественных фильмов. А вот для сети из сериалов зависимость практически не изменяется с ростом степени, тем самым позволяя отнести ее к некоррелированным.

Следует заметить, что зависимость средней степени ближайших соседей вершины объединенной сети Мос-и Ленфильма и сети им. М.Горького основных актеров, изученных в работе [3], при $k > 100$ практически не зависела от k и являлась некоррелированной.

3. Избыточность и локальная кластеризация

В 1992 году американский социолог Рональд Стюард Бёрт, исследуя явления социальных сетей, ввел понятие структурных дыр (structural holes) [7]. *Структурная дыра* – это отсутствие связи между определенными частями сети. Если рассматривать эффективное распространение информации или распределения трафика по сети, то такие дыры очень плохи для сети, так как они уменьшают альтернативные пути информации, которые могут быть выбраны. С другой стороны, они очень полезны для центрального узла, друзьям которого недостает связей. Они дают вершине власть над потоком информации между этими друзьями. Если два друга вершины не связаны напрямую, и информация друг о друге проходит через взаимные связи с данной вершиной, то вершина может управлять потоком этой информации. Для изучения существования структурных дыр Р. Бёрт ввел понятие *избыточности*.

Избыточность вершины R_i означает среднее число связей узла ко всем его соседям. Минимальное число избыточности равняется 0, максимальное $k_i - 1$.

$$R_i = \frac{2 \cdot \sum_{j,m=1}^N v_{ij} v_{im} \cdot v_{jm}}{k_i},$$

где $\sum_{j,m=1}^N v_{ij} v_{im} \cdot v_{jm}$ обозначает количество связей между соседями вершины.

Для безмасштабной сети следует ожидать зависимость вида

$$R(k) \sim k^\gamma.$$

На рис. 6 изображен график зависимости избыточности вершины от ее степени. График подтверждает такой характер зависимости только для полной сети и сети, состоящей из сериалов, имеющих практически одинаковый показатель степени, разница между значениями которого составляет 0,05 вне зависимости от учета эпизодических актеров. Сеть из художественных фильмов степенной зависимости не подчиняется и является просто возрастающей функцией.

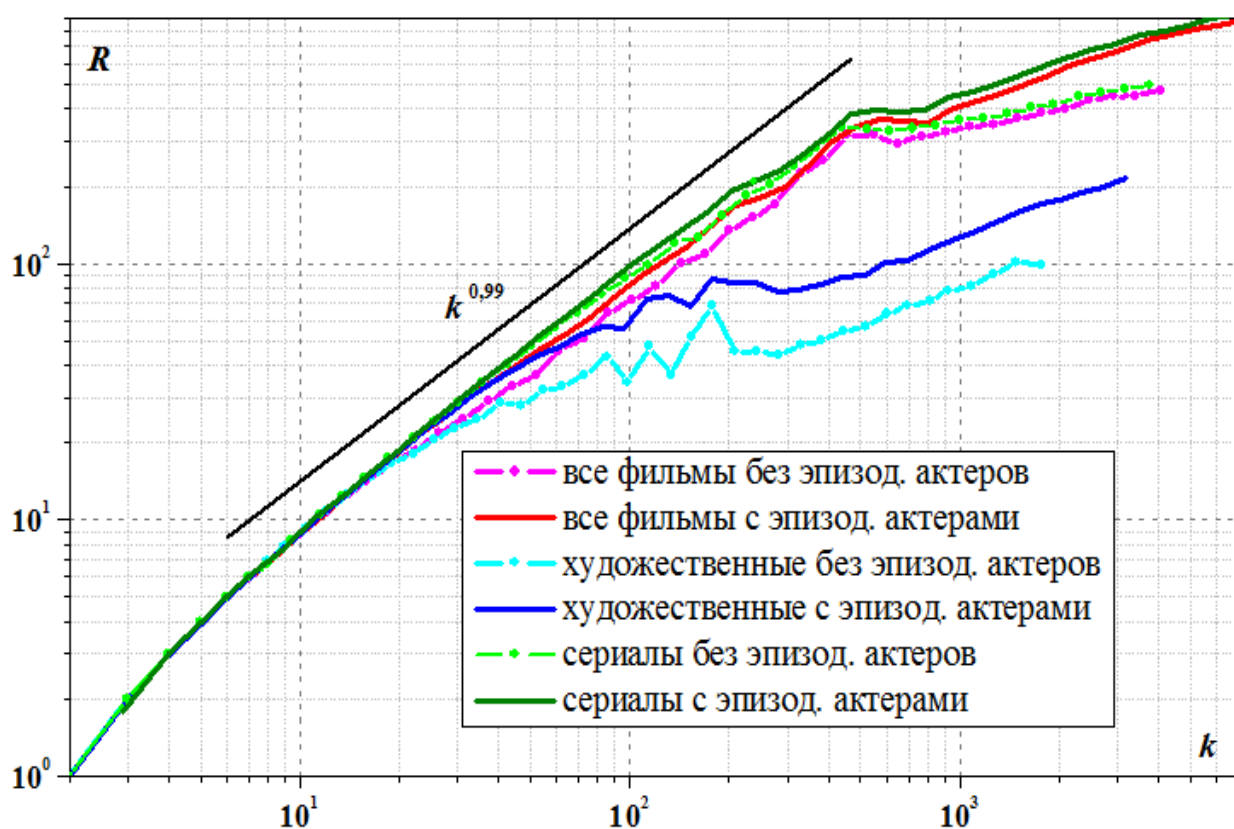


Рис. 6. Избыточность вершины в зависимости от ее степени

В области значений степени шириной чуть менее двух порядков зависимость характеризуется показателем $\gamma \approx 0,94$ для полной сети актеров с учетом эпизодических актеров и $\gamma \approx 0,92$ без учета эпизодических актеров; $\gamma \approx 0,99$ с учетом эпизодических актеров и $\gamma \approx 0,97$ для сериалов без учета эпизодических актеров. Для художественных фильмов зависимость не описывается степенным видом.

Обращают на себя внимание провалы в районе значений $k \approx 100$, 177 для художественных фильмов и $k \approx 450$ для сериалов и полной сети. При рассмотрении зависимости на рис. 7 выпадает намного крупнее.

Очевидно, что величина R тесно связана с *локальным коэффициентом кластеризации* (the local clustering coefficient), введенным Дунканом Уоттсом и Стивенем Страттсом в 1998 г. [8]. Определяется как средняя доля пар соседей вершины, которые являются также соседями друг друга. Применительно к ки-

ноактерам величина коэффициента кластеризации говорит о том, как тесно связаны режиссёрские вкусы, ставящие одних и тех же актёров в свои фильмы.

Локальный коэффициент кластеризации интересен по многим причинам. Во-первых, доказано, что во многих сетях имеет место зависимость от степени: вершины с более высокой степенью имеют более низкий коэффициент кластеризации в среднем. Во-вторых, локальный коэффициент кластеризации также можно использовать для исследования существования «структурных дыр».

Для вершины i коэффициент кластеризации

$$C_i = \frac{R_i}{k_i - 1}.$$

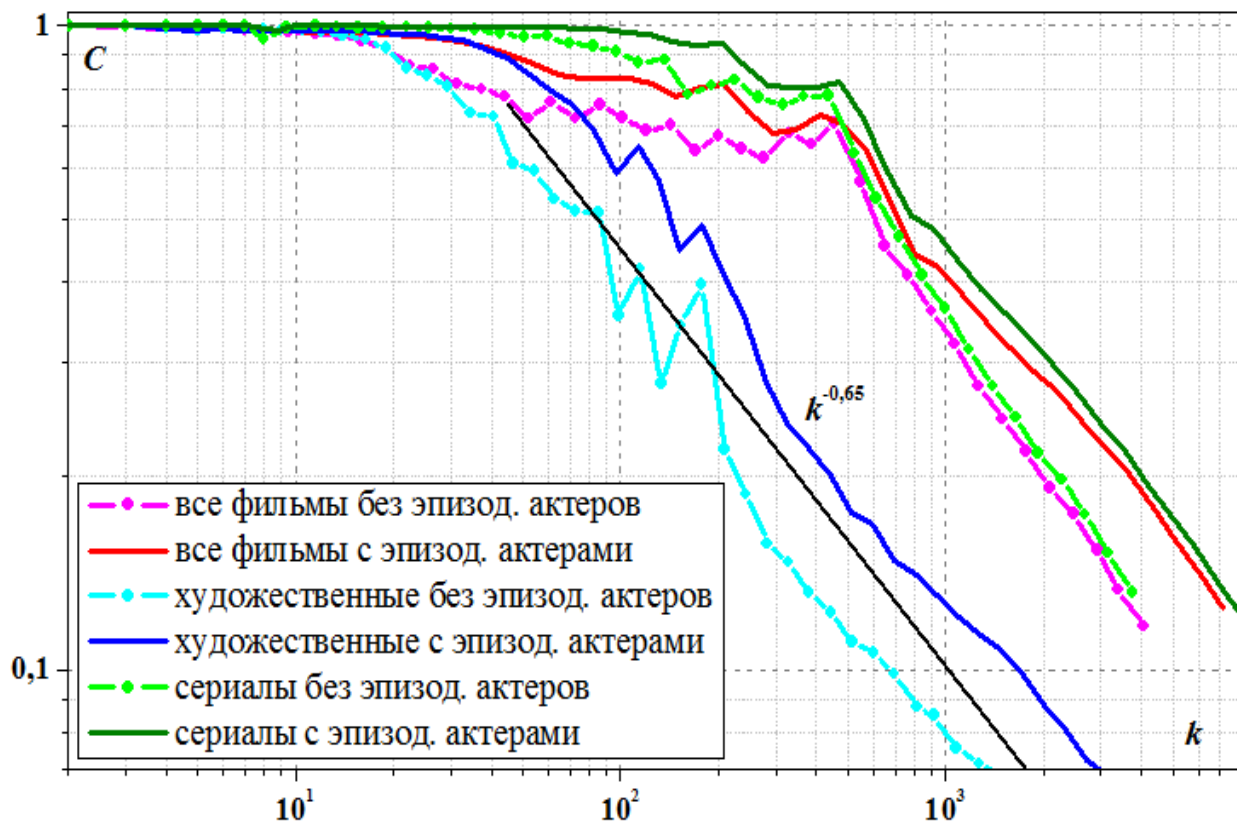


Рис. 7. Коэффициент кластеризации вершины в зависимости от ее степени

Область промежуточной асимптотики с показателем $\gamma \approx 0,68$ имеет ширину около двух порядков для художественных фильмов без учета эпизодических актёров и $\gamma \approx 0,65$ с учетом эпизодических актёров. Для полной базы данных и сериалов степенная зависимость характерна лишь в хвосте. Добавление эпизодических актёров практически не влияет на изменения в показателях степени.

Зная C_i для каждой вершины, можно определить *глобальный коэффициент кластеризации*, который определяется как

$$\langle C \rangle = \sum_{i=1}^N C_i.$$

Значения глобального коэффициента кластеризации для киноактеров полной базы составляют с учетом эпизодических актеров 79,0%, без учета – 77,7 %, для художественных фильмов с учетом эпизодических актеров 78,7%, без учета – 77,4 % и для сериалов с учетом эпизодических актеров 78,2 %, без учета – 76,6 %. Обращают на себя внимания очень высокие значения глобальных коэффициентов кластеризации. Для сравнения, результат, полученный для Живого журнала [1], равный 31 % для взаимных связей, на 48 % ниже, чем для полной базы фильмов с учетом эпизодических актеров. Высокие значения коэффициента кластеризации связаны с особенностями структуры образования связей в сетях киноактеров.

Рис. 7 демонстрирует общую тенденцию убывания коэффициента кластеризации соседей вершины по мере увеличения ее степени. Степенной зависимостью $C(k) \sim k^{-\gamma}$ описывается только сеть, состоящая из художественных фильмов. Для полной базы фильмов и сериалов зависимость разбивается на два степенных участка. Показатель степени первого участка для полной базы данных $\gamma \approx 0,13$ без учета эпизодических актеров и $\gamma \approx 0,10$ с учетом эпизодических актеров и второго участка, выражающего в виде степенного хвоста $\gamma \approx 0,78$ без учета эпизодических актеров и $\gamma \approx 0,6$ с учетом эпизодических актеров. Исходя из соображений, что показатели степеней первого участка достаточно малы, в дальнейшем ими можно пренебречь.

В данном случае, как и для величины средней степени ближайших соседей, влияние эпизодических актеров практически не сказывается на изменении степенной зависимости $R(k)$ и, тем самым, не отражается на статистических свойствах описанных выше сетей. Что касается зависимости $C(k)$, то при $k > 450$ она становится более крутой для полной сети и сериалов без учета эпизодических актеров.

Следует заметить, что в работе [3] $C(k) \sim k^{-1}$, что на 0,35 выше результатов для художественных фильмов без учета эпизодических актеров.

4. *Кратчайшие пути и малые миры*

Одним из самых поразительных и широко обсуждающихся в литературе феноменов сети является *эффект малого мира*, состоящий в том, что во многих, а возможно, и в большинстве сетей типичное расстояние между вершинами удивительно мало. Первым эмпирическим доказательством «явления тесного мира» считается эксперимент С. Милгрэма, который показал, что среднее «расстояние» между любыми парами людей, в сообществе, которое он исследовал, приблизительно равно шести [8].

Наименьшее число связей, необходимое для перехода от узла i к узлу j , называется *расстоянием* d_{ij} между двумя вершинами i, j (длины связей между смежными вершинами считаются равными единице). Максимальное значение d_{ij} называется диаметром графа. Если число вершин в графе велико (для пол-

ной базы данных, включая эпизодических актеров, свыше 106 тыс., исключая – 60 тыс.), а расстояния между ними – нет, то говорят об эффекте *малого мира*.

На рис. 8 показано распределение $U(d)$ расстояний между случайно выбранными вершинами для сетей киноактеров с учетом и без учета эпизодических актеров. График имеет острый пик в районе значения $d=3$ для художе-

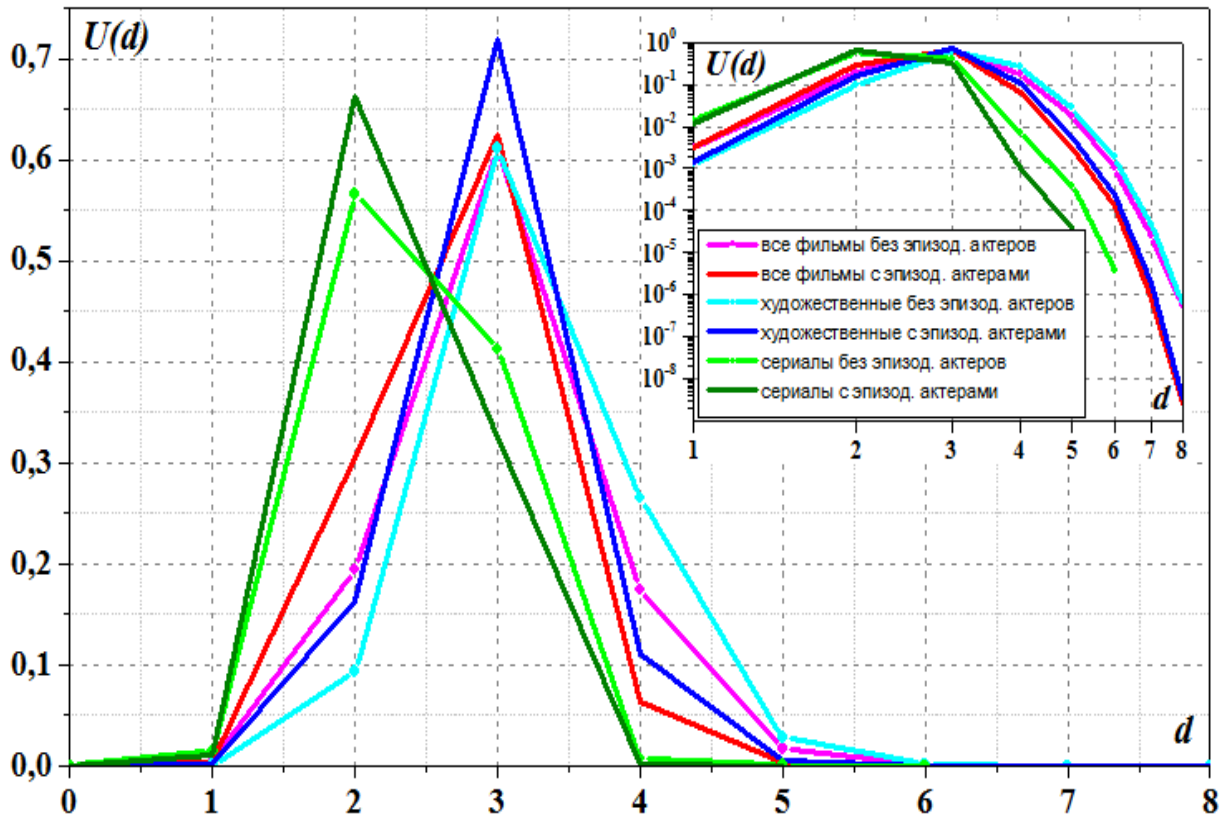


Рис. 8. Распределение пар вершин по кратчайшему расстоянию между ними

Плотность нормирована на вероятность того, что между вершинами существует путь, составляющую для полной сети с учетом эпизодических актеров 99,4 %, без учета – 98,6 %, для художественных фильмов с учетом эпизодических актеров 99,3 %, без учета – 98,1 %, для сериалов с учетом эпизодических актеров 99,96 %, без учета – 99,75 %.

На врезке построены те же графики в полулогарифмическом масштабе.

ственных фильмов и полной сети и $d=2$ для сериалов. Среднее расстояние между вершинами сети для полной сети и для художественных фильмов $\langle d \rangle = 4,5$, для сериалов $\langle d \rangle = 3$. Данный график не подтверждает концепцию шести степеней разделения С. Милгрэма. Актёрский мир оказывается более тесным, чем просто людское сообщество. Влияние эпизодических актеров на распределение $U(d)$ не распространяется на статистические свойства сети.

ДИНАМИКА СЕТИ СОТРУДНИЧЕСТВА КИНОАКТЕРОВ

Большой интерес представляет исследование динамики числа выпущенных фильмов для различных выборок (см. рис. 9). Следует уточнить, что если съемки кинофильмов длились более одного года, то при учете данных считает-

ся год выпуска. Всего таких фильмов в базе 275 штук. Как видно, для сериалов в период с первого выпуска (1969 г.) по 1997 г. количество выпущенных фильмов в год практически не меняется, а с 1998 г. происходит резкое увеличение числа выпущенных сериалов за год.

Что касается динамики числа выпущенных фильмов для полной базы данных и для художественных фильмов, то в 1921 г. съемки фильмов сходят на нет (только 9 фильмов), т.к. начало 20-х годов (1921-1923 гг.) прошлого века в России – это период гражданской войны, голода, разрухи, экономического хаоса и разгула бандитизма. 1928 год был отмечен рядом событий, оставивших заметный след в истории советского изобразительного искусства. Как видно из графика, в этот год отмечался и большим количеством снятых фильмов (115 шт.). С 1951 г. начинается развитие советского кинематографа вплоть до 1991 г. Обращает на себя внимание некоторый провал в графике, приходящийся на первую половину восьмидесятых годов прошлого века.

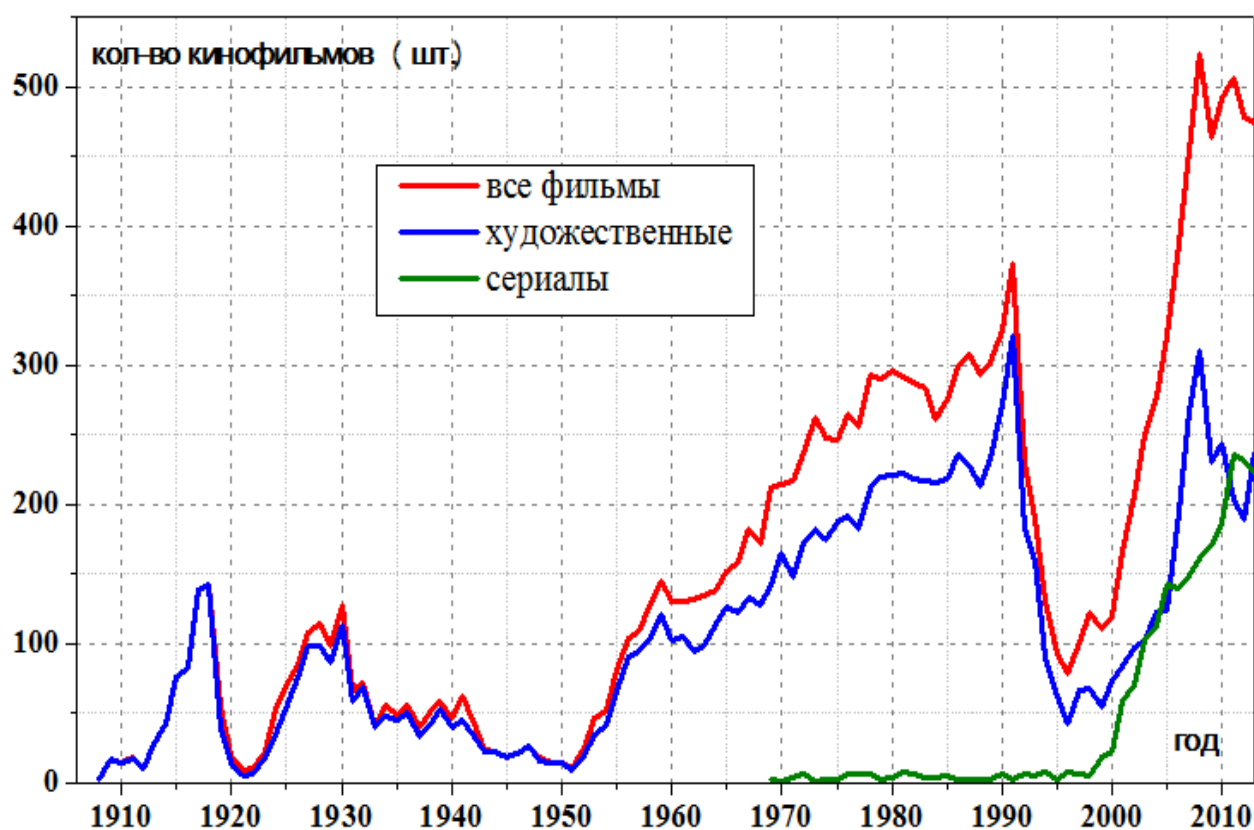


Рис. 9. Динамика числа выпущенных фильмов в период с 1908 по 2013 г.

После распада СССР (конец 1991 г.) и вплоть до 1996 года происходит снижение количества съемок фильмов. С 1997 г. по настоящее время происходит возрождение кинематографа, исключая период 1999 г. (дефолта).

На рис.10-11 изображено изменение размера сети и числа ребер со временем. Как видно из рис. 10, за последние 15 лет с 1998 г. (6846 человек с учетом

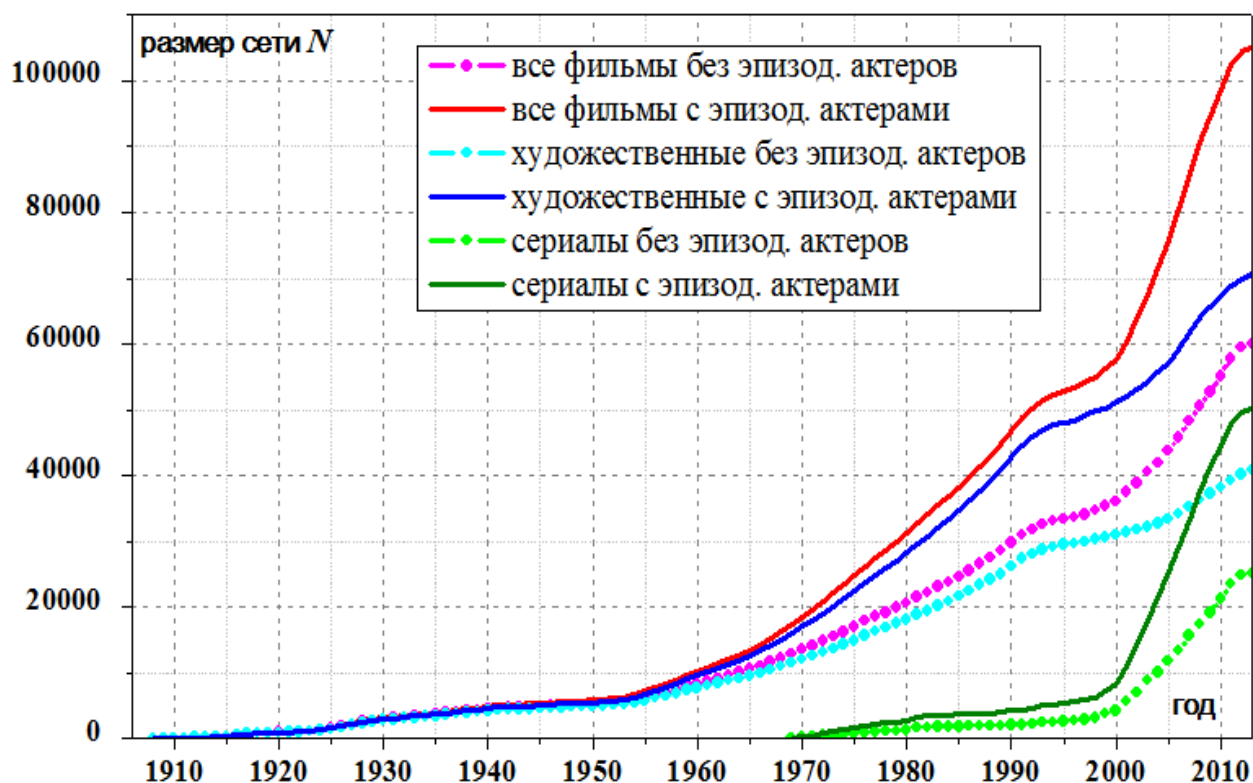


Рис. 10. Изменение размера сети киноактеров со временем

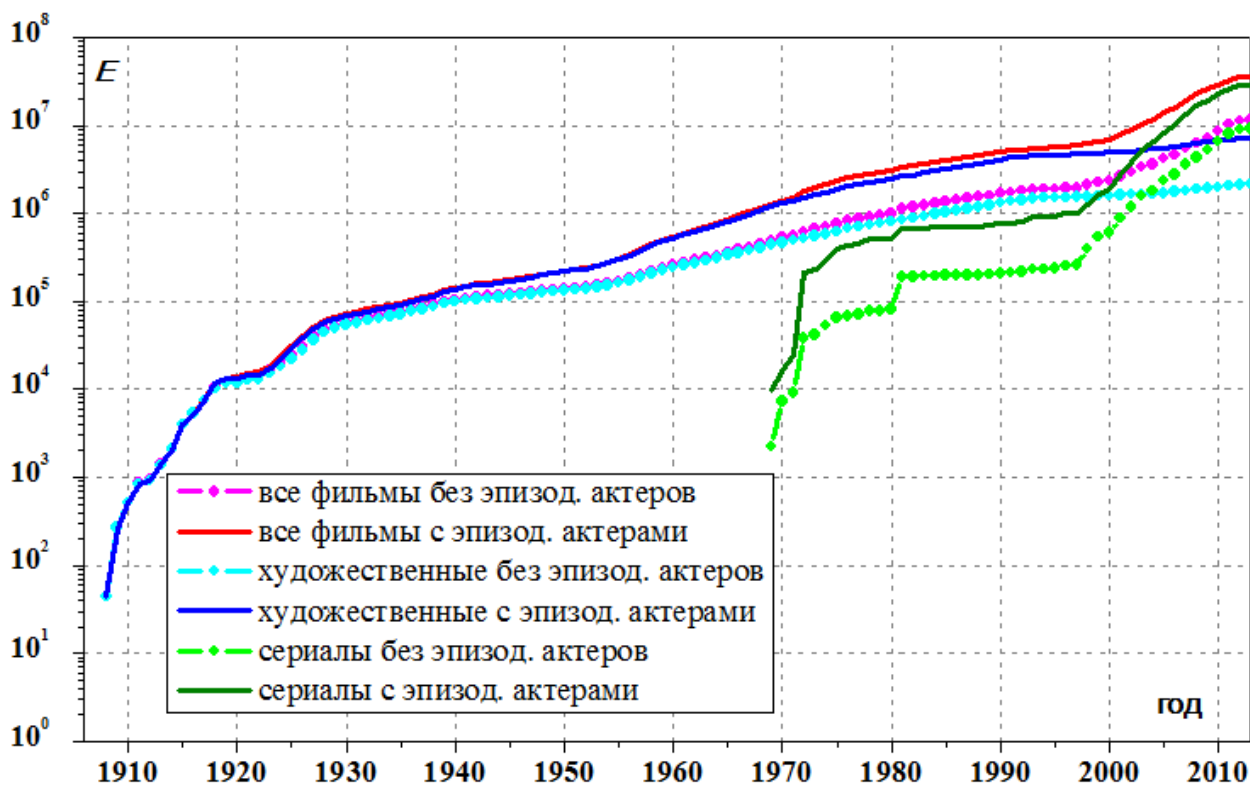


Рис. 11. Изменение числа ребер сети со временем в полулогарифмическом масштабе

эпизодических актеров, без учета – 3164 актера) по настоящее время размер сети для данных, состоящих из сериалов, вырос в 8 раз и составил 87,4 % от общего числа вершин. Размер сети для художественных фильмов за период с 1956 по 1991 г. увеличился с учетом эпизодических актеров на 53 %, без учета – на 52 %, а с 1992 г. по настоящее время – на 35% (с учетом эпизодических актеров),

на 31 % без учета эпизодических актеров. Размер сети полной базы за период с 1956 по 1991 год вырос на 39 % с учетом эпизодических актеров, на 41 % без учета, а с 1992 года по настоящее время – на 52 % с учетом эпизодических актеров, на 47 % без учета. Что касается изменения числа ребер со временем, то для сериалов за 15 лет общее число с учетом эпизодических актеров выросло в 22,6 раза, без учета – в 23,4 раза, для художественных фильмов с учетом эпизодических актеров в 1,5 раза, а для полной базы с учетом эпизодических актеров – в 5,7 раза, без учета – в 5,4.

На рис. 12 показана зависимость изменения числа ребер сети от ее размера. Как видно из рисунка, рост описанных сетей происходит степенным образом $E(N) \sim N^\gamma$. Следует отметить, что добавление эпизодических актеров к

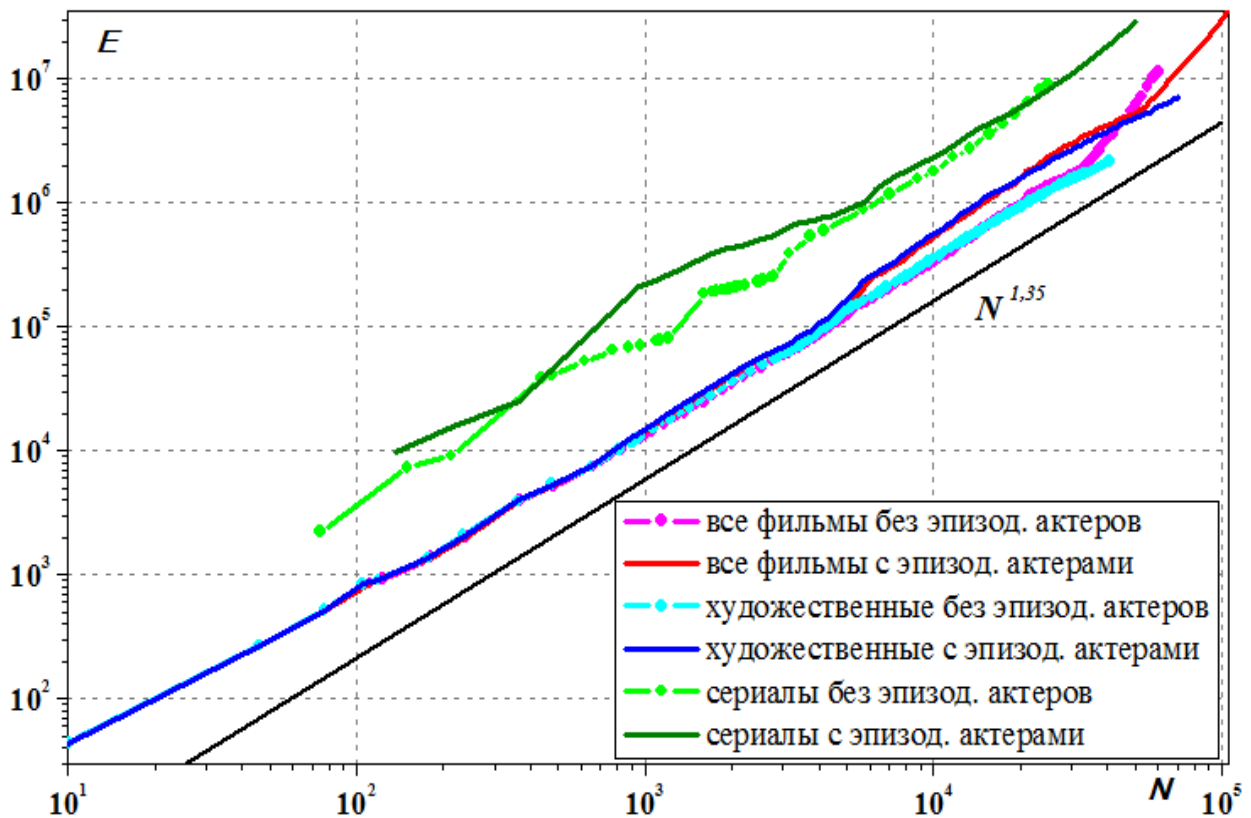


Рис. 12. Изменение числа ребер сети от ее размера в логарифмическом масштабе

Рост числа ребер сети от ее размера описывается степенным законом с $\gamma = 1,3$ с учетом эпизодических актеров, $\gamma = 1,38$ – без учета эпизодических актеров для сериалов, $\gamma = 1,44$ с учетом эпизодических актеров, $\gamma = 1,35$ – без учета эпизодических актеров для художественных фильмов и $\gamma = 1,5$ с учетом эпизодических актеров, $\gamma = 1,43$ без учета эпизодических актеров для полной базы фильмов.

любой выборке сети практически не влияет на значения показателя степени, разница которых составляет меньше 0,1.

Выводы

Исследование статистических свойств сети сотрудничества киноактеров российского кинематографа продемонстрировало масштабно-инвариантный характер сети, выражающийся в степенном виде распределения вершин по числу связей, зависимостей избыточности вершины от ее степени для полной сети и сети из сериалов и кластеризации вторичных связей.

Показатель распределения вершин по степени по полной базе фильмов с учетом эпизодических актеров российского кинематографа отличается от голливудской базы данных на 0,92. Разница между показателем сети, состоящей только из художественных фильмов без учета эпизодических актеров, и объединенной сети, Мос- и Ленфильма, исследованной в работе [3], составляет 0,26. Исследование влияния эпизодических актеров в сети продемонстрировало увеличение показателя в распределении $u(k)$. Выявлены некоторые расхождения в значениях показателей зависимостей средней степени ближайших соседей от степени вершины и их кластеризации с работой [3]. Сеть киноактеров РФ, за исключением выборки по сериалам, так же, как и сеть киноактеров Голливуда, является ассортативной, в то время как для объединенной сети Мос- и Ленфильма присутствие коррелированных свойств не выявляется. Кластеризация вторичных связей для сети, исследованной в данной работе, оказывается больше на 0,35, чем для объединенной сети Мос- и Ленфильма.

Плотность распределения кратчайших путей между случайными парами вершин характеризуются очень резким пиком, расположенным в районе тройки для полной базы фильмов и сети, состоящей их художественных фильмов, и в районе двойки для сети из сериалов, что оказывается даже в два раза меньше, чем в концепции С. Милгрэма о шести степенях разделения. Сети сотрудничества киноактеров РФ являются малыми мирами, и актеры находятся в среднем в 4,5 звеньях друг от друга для полной базы фильмов и художественных, и в 3 звеньях – для сериалов.

Продемонстрирована динамика сети сотрудничества киноактеров, которая показывает масштабно-инвариантный характер в изменении числа ребер сети в зависимости от ее размера.

Авторы выражают благодарности программисту сайта www.kino-teatr.ru Петру Кирееву за предоставленные данные и Андрею Подлазову за полезное обсуждение работы.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Митин Н.А., Подлазов А.В., Щетинина Д.П. Исследование сетевых свойств Живого журнала. // Препринты ИПМ им. М.В. Келдыша. 2012. № 78. 16 с. URL: <http://library.keldysh.ru/preprint.asp?id=2012-78>
2. Albert R., Barabási A.-L. Statistical mechanics of complex networks. *Reviews of Modern Physics* 74, 47-97 (2002).
3. Гаджиев Б.Р., Прогулова Т.Б., Щетинина Д.П., Кузин А.Е. Эволюция сети киноактеров // «Математика. Компьютер. Образование». Сб. трудов XVI международной конференции. Под общей редакцией Г.Ю. Ризниченко. — Ижевск: Научно-издательский центр «Регулярная и хаотическая динамика». — 2009. — Т.1. — С. 398-410.
4. Chang Pei-Pei, Chen Kan, He Yue, etc. Model and empirical study on some collaboration networks, *Physica A* 359 (2006), p.835-852.
5. 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.
6. Pator-Satorras R., Vazquez A. and Vespignani A. *Phys.Rev.Lett.* 87 (2001), p/258701.
7. Burt R. S. *Structural Holes: The Social Structure of Competition*, Harvard University Press, Cambridge, MA (1992).
8. Watts D.J. and Strogatz S.H. *Nature* 393 (1998), p.440.