

ЭВОЛЮЦИЯ СЕТИ КИНОАКТЕРОВ

Гаджиев Б. Р., Прогулова Т. Б., Щетинина Д. П., Кузин А. Е.

В работе на основе данных по художественным фильмам киностудий России была построена сеть сотрудничества киноактеров. Проведен анализ данных, построены распределение степеней, зависимость коэффициента кластеризации от степени узла и определена зависимость средней степени ближайших соседей вершин от их степени. Предложена модель эволюционирующей сети, позволяющая генерировать сети, статистически эквивалентные сети киноактеров. Из сравнения теоретических результатов и результатов обработки данных определены параметры модели.

Введение. Общая теория систем начинается с определения «системы» как комплекса взаимодействующих компонентов, формирующих организованное целое. Такое определение естественным образом приводит к удобному и эффективному моделированию самых разнообразных систем (физических, как природных, так и искусственных, биологических, социальных) на языке теории графов (сетей), когда компоненты — суть вершины, а их взаимодействия — связывающие эти вершины ребра. В последние годы применение методов статистической механики сложных сетей позволило описать многие сложные системы реального мира, представимые в виде сетей, такие как Интернет, WWW, сети цитирований, различные сети соавторства научных публикаций и т.д. (Albert and Barabasi, 2002; Dorogovtsev and Mendes, 2002).

Барабаси и Альберт (Albert and Barabasi, 2000) исследовали сеть сотрудничества киноактеров на основе Internet Movie Database. В этой сети каждый актер — узел, и два актера связаны, если они в течение своей карьеры хотя бы раз снимались в одном и том же кинофильме. Исследуемая сеть содержала 212 250 вершин–актеров, и 3 045 787 связей. Средняя степень вершины $\langle k \rangle = 28.78$. Распределение степеней при больших значениях k описывается степенным законом $p(k) \sim k^{-\gamma}$ с показателем $\gamma = 2.3$.

Сеть киноактеров растет непрерывно добавлением новых актеров (вершины). Однако, важный вклад в рост сети дают другие часто происходящие события: старый актер играет в новом фильме, устанавливая новые внутренние связи с актерами (вершинами), с которыми он или она прежде не играл.

Для описания эволюции сети киноактеров Барабаси и Альберт предложили так называемую расширенную модель. В этой модели развитие сети происходит следующим образом. Процесс роста начинается с m_0 изолированных вершин, и на каждом временном шаге выполняется одна из следующих двух операций:

(i) С вероятностью p мы добавляем m ($m \leq m_0$) новых связей, для каждой из которых одна из вершин выбирается случайным образом, а другая с вероятностью

$$P(k_i) = \frac{k_i + 1}{\sum_j (k_j + 1)},$$

то есть новая связь «предпочтительно» связывается с вершиной более высокой степени.

(ii) С вероятностью $(1-p)$ мы добавляем новую вершину с m новыми связями. Связывание с узлом i , уже присутствующим в системе, осуществляется с вероятностью $P(k_i)$.

Сравнение результатов обработки данных для сети актеров Голливуда с результатами моделирования показало, что для этой сети $p = 0.937$ и $m = 1$. Это означает, что эволюция сети происходит в основном не за счет добавления в сеть новых актеров, а за счет возникновения новых связей между уже присутствующими в сети актерами.

Анализ алгоритма расширенной модели Барабаси показывает, что такие процессы описываются распределением Цаллиса, для которого характерно плато при небольших значениях степеней и степенное распределение при больших значениях k (Гаджиев и др., 2008; Progulova and Gadjiev, 2008). Таким образом, эволюция сети актеров Голливуда хорошо описывается распределением Цаллиса (Tsallis, 1999).

В работе на основе данных по художественным фильмам киностудий России мы исследовали сеть сотрудничества киноактеров. Были построены распределение степеней и зависимость коэффициента класте-

ризации от степени узла. Для выяснения характера корреляций сети определена зависимость средней степени ближайших соседей вершин от их степени. На основе предложенной модели выявлены механизмы роста сети актеров.

Анализ данных. Для построения сети сотрудничества киноактеров мы использовали данные, размещенные на официальных сайтах киностудий Ленфильм (www.lenfilm.ru) и Мосфильм (www.mosfilm.ru). В период с 1918 по 1992 год эти киностудии — наиболее крупные производители фильмов в России — выпустили 2910 художественных фильма. Необходимо отметить, что данные на сайтах киностудий содержат списки только основных актеров, снимавшихся в ролях первого и второго плана.

Мы строили сеть сотрудничества киноактеров следующим образом: вершинами являются актеры, две вершины соединены ребром, если соответствующие актеры в течение своей карьеры хотя бы раз снимались в одном и том же фильме. В построенной сети содержится 6997 вершин (актеров) и 75940 ребер. Для анализа динамики роста сети мы строили зависимости от времени числа выпускаемых фильмов (рис. 1), общего числа снявшихся актеров (рис. 2) и числа ребер, т.е. сотрудничеств между актерами (рис. 3). Также, имея в виду рост сети со временем, была построена зависимость числа ребер от числа вершин (рис. 4).

Было построено распределение степеней (рис. 5). Максимум распределения соответствует $k = 6$. При $k > 6$ распределение хорошо описывается степенным законом: $p(k) \sim k^{-\gamma}$. Мы определяли значение показателя, используя метод максимального правдоподобия: $\gamma = 2.14$.

Для анализа локальной структуры сети мы вычисляли значения коэффициентов кластеризации вершин и строили зависимость коэффициента кластеризации от степени k вершины (рис. 6). Среднее значение коэффициента кластеризации — $\langle c \rangle = 0.76$.

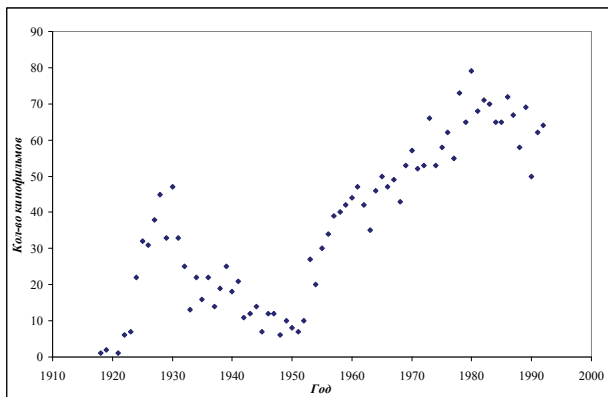


Рис. 1. Зависимость от времени числа кинофильмов, выпускаемых за год.

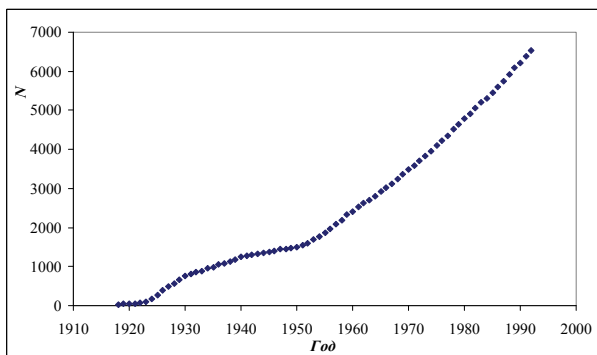


Рис. 2. Изменение размера N сети актеров со временем.

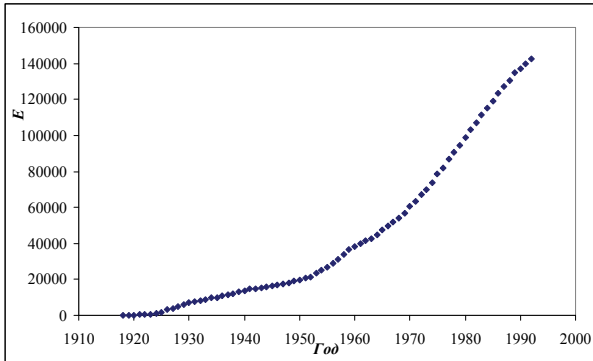


Рис. 3. Изменение числа ребер E сети актеров со временем.

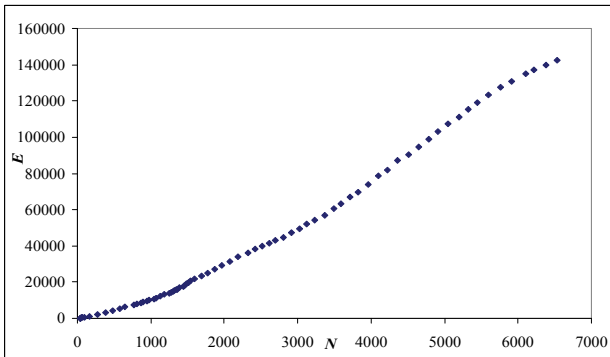


Рис. 4. Зависимость числа ребер E сети актеров от размера сети N .

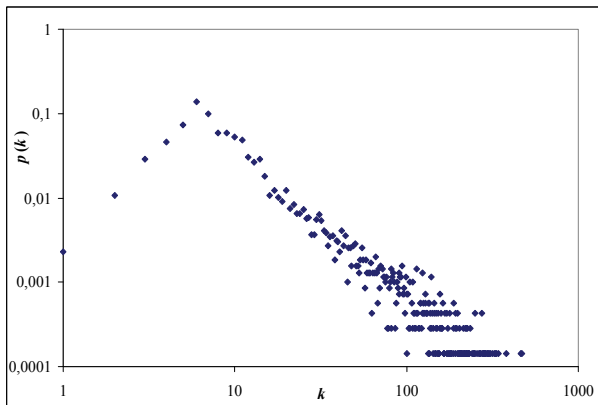


Рис. 5. Распределение степеней для сети сотрудничества киноактеров.

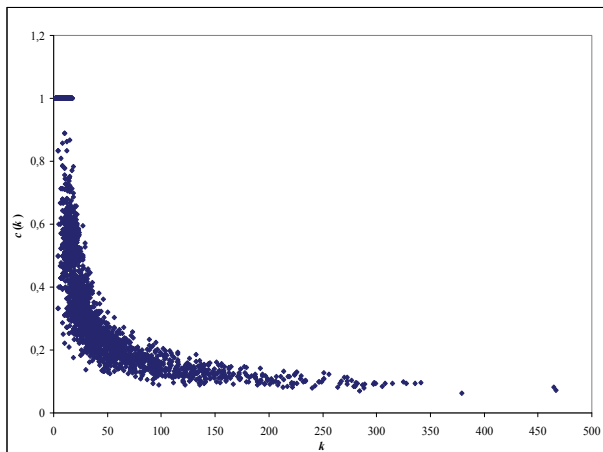


Рис. 6. Зависимость коэффициента кластеризации вершины от ее степени для сети сотрудничества киноактеров.

Для определения характера корреляций степеней вершин в сети была построена зависимость средней степени ближайших соседей вершины от ее степени (рис. 7). Форма зависимости показывает, что при

$k > 100$ функция k_{nn} практически постоянна и, следовательно, сеть является некоррелированной.

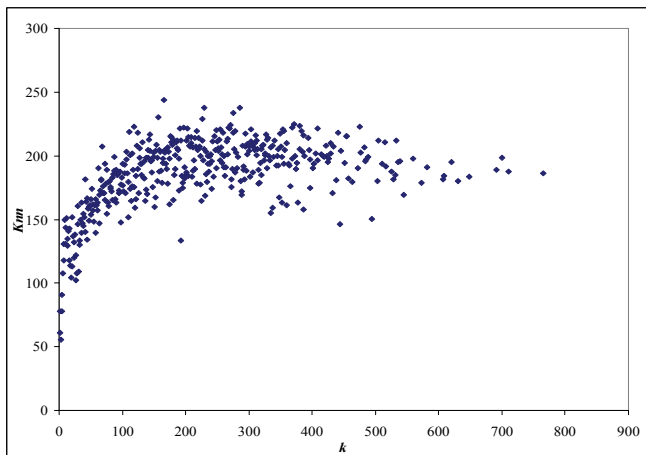


Рис. 7. Зависимость средней степени ближайших соседей k_{nn} вершины от ее степени k для сети сотрудничества киноактеров.

Отдельно мы анализировали сеть сотрудничества актеров Центральной киностудии детских и юношеских фильмов имени М. Горького. На рис. 8–10 представлены распределения степеней, зависимость коэффициента кластеризации и средней степени ближайших соседей k_{nn} вершины от ее степени k .

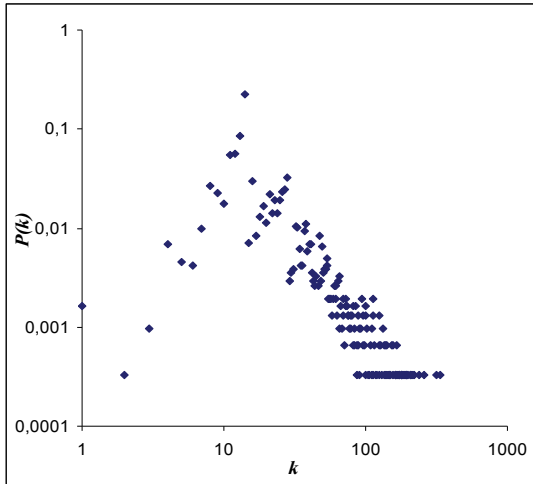


Рис. 8. Распределение степеней для сети сотрудничества актеров киностудии им. Горького.

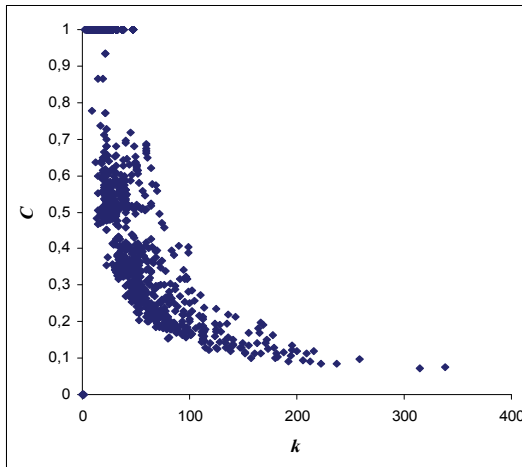


Рис. 9. Зависимость коэффициента кластеризации вершины от ее степени для сети сотрудничества актеров киностудии им. Горького.

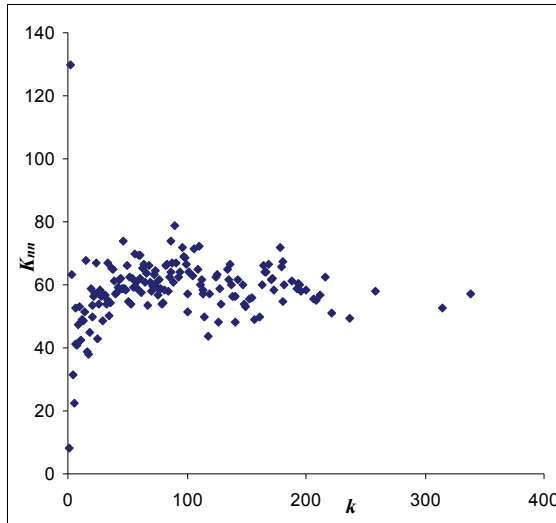


Рис. 10. Зависимость средней степени ближайших соседей k_{nn} вершины от ее степени k для сети сотрудничества киноактеров.

Как и в предыдущем случае, распределение степеней имеет ярко выраженный максимум, зависимость коэффициента кластеризации от степени $c(k) \sim k^{-1}$ и k_{nn} практически не зависит от k .

Моделирование. В рамках принципа максимума энтропии вывод распределения Цаллиса предполагает два ограничения — нормировку распределения вероятностей и ограничение на среднее значение. Это приводит к распределению Цаллиса в виде

$$P(k) = \frac{1}{Z} \left(1 - (1-q) \frac{k}{k_0} \right)^{\frac{1}{1-q}},$$

где k_0 — средняя степень, q — параметр нежесткости, Z — статистическая сумма (Tsallis, 1999). Как отмечалось выше, это распределение достаточно хорошо описывает распределение степеней сети актеров Голливуда. При небольших значениях степеней распределение Цаллиса

имеет плато, а при больших значения k зависимость приобретает степенной вид. Все полученные нами из анализа данных распределения степеней для изучаемых сетей киноактеров не имеют плато при небольших значениях k , а имеют явно выраженный максимум, хотя при больших значениях степеней они также имеют степенной характер. В рамках метода максимума энтропии такой вид распределения можно получить, если ввести дополнительное ограничение на момент второго порядка. В этом случае мы получаем q -экспоненциальное распределение в виде:

$$P(k) = \frac{1}{Z} (1 - \xi(1 - q)(k - \eta)^2)^{\frac{q}{1-q}}. \quad (1)$$

При этом параметр η определяет положение максимума зависимости $p(k)$. Таким образом, распределение (1) при $k = \eta$ имеет максимум, а при больших значениях k имеет степенную форму. Распределение (1) достаточно хорошо описывает распределение степеней для маломировых сетей (small-world network) Ваттса–Строгатца, для которых характерны большие значения коэффициентов кластеризации (Hasegawa, 2006). Алгоритм их генерации предполагает, что изначально сеть представляет собой регулярный граф. В последующие моменты времени происходят пересвязывания ребер, что порождает неоднородную сеть с большой средней кластеризацией.

Маломировым свойством обладают также сверх мало-мировые сети (smallest-world network), которые сохраняют регулярность исходного графа, а эволюционируют за счет добавления очень небольшого числа новых узлов, которые со временем соединяясь с все большим числом узлов сети, приобретают очень большие степени (Dorogovtsev and Mendes, 2002).

Начальный этап развития сети актеров можно промоделировать с помощью следующей модификации алгоритма обобщенной модели Барабаси. Мы, как и в случае обобщенной модели Барабаси, начинаем с m_0 изолированных вершин, и на каждом временном шаге с вероятностью p добавляем m ($m \leq m_0$) новых ребер, и с вероятностью $(1 - p)$ добавляем новую вершину с m новыми связями. Наша модификация состоит в том, что величина m случайна и распределена в соответствии с некоторым законом $P(m)$ на интервале от 1 до m_{max} .

На рис. 11 представлено распределение степеней для сгенерированной в соответствии с данной моделью сети.

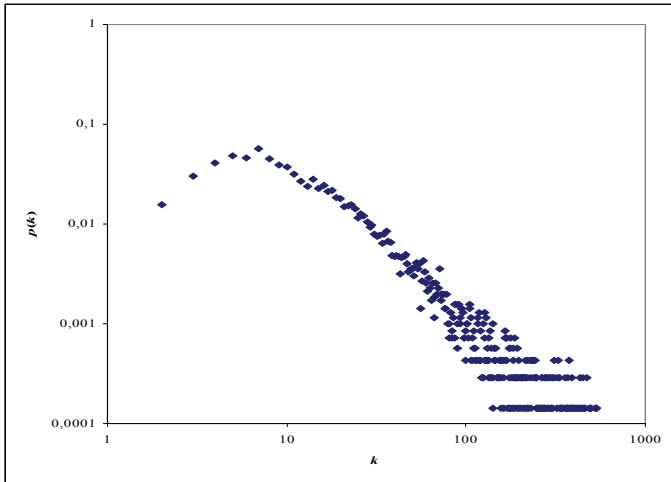


Рис. 11. Распределение степеней для сети, сгенерированной с помощью модифицированного алгоритма ($m_0 = 7$, $N = 7000$, $p = 0.5$, m равномерно распределено на интервале от 2 до 8).

Подгонка результатов обработки данных сетей киноактеров с помощью формулы (1) показывает, что значение параметра неэкстенсивности $q = 0.51$. Сравнение результатов обработки данных с результатами моделирования показывает, что сеть киноактеров России также в основном растет за счет добавления новых связей ($p = 0.91$).

Заключение. В работе на основе данных по художественным фильмам киностудий России исследовалась сеть сотрудничества киноактеров.

Мы анализировали механизмы роста сети актеров. Очевидно, сеть непрерывно растет добавлением новых узлов (актеров), которые связываются с уже существующими в сети узлами. Однако, если в фильме снимаются актеры, уже присутствующие в сети, но раньше совместно не снимавшиеся ни в одном фильме, то в сеть добавляются новые связи. Моделирование этих процессов приводит к распределению Цаллиса,

которое достаточно хорошо описывает сеть актеров Голливуда. Мы показываем, что полученное нами распределение степеней сети сотрудничества киноактеров России этим распределением не описывается. Отметим, что в рамках принципа максимума энтропии вывод распределения Цаллиса предполагает два ограничения — нормировку распределения вероятностей и ограничение на среднее значение. Мы включаем еще одно ограничение — на дисперсию, выводим соответствующее q -экспоненциальное распределение и, используя метод максимального правдоподобия, показываем, что оно достаточно хорошо описывает распределение степеней построенной сети киноактеров. В работе предлагается микроскопическая модель, являющаяся модификацией обобщенной модели Барабаси, в которой число поступающих в сеть ребер описывается некоторым распределением. Сравнение результатов моделирования с результатами анализа данных позволяет заключить, что сеть киноактеров растет в основном за счет добавления новых связей между вершинами, уже присутствующими в сети и что этой сети присущи особенности сверх мало-мировых сетей.

СПИСОК ЛИТЕРАТУРЫ

- Гаджиев Б.Р., Калинкина Е.А., Прогулова Т.Б. Самоорганизация сети слов // Четвертые Курдюмовские юбилейные чтения: Материалы конференции / Международная междисциплинарная научная конференция «Синергетика в естественных науках». — Тверь: Твер. гос. ун-т, 2008. — С. 322–326.
- Albert R., Barabási A.-L. Topology of evolving networks: Local events and universality // *Phys. Rev. Lett.* — 2000. — Vol. 85. — P. 5234–5237.
- Albert R., Barabási A.-L. Statistical Mechanics of Complex Networks // *Rev. Mod. Phys.* — 2002. — Vol. 74. — P. 43–97.
- Dorogovtsev S.N., Mendes J.F.F. Evolution of networks // *Adv. Phys.* — 2002. — Vol. 51. — P. 1079–1187.
- Gadjiev B.R., Progulova T.B. Topology properties of written human language // International Conference in Statistical Physics, Kolympari — Greece 14–18 uly 2008. — 2008. — P. 93.
- Tsallis C. Nonextensive statistics: theoretical, experimental and computational evidences and connections // *Braz. J. Phys.* — 1999. — Vol. 29. — P. 1–35.
- Hasegawa H. Nonextensive aspects of small-world networks // *Physica A: Statistical Mechanics and its Applications.* — 2006. — Vol. 365. — P. 383–401.

EVOLUTION OF THE MOVIE ACTOR COLLABORATION NETWORK

Gadjiev B. R., Progulova T. B., Shchetinina D. P., Kuzin A. E.

We represent model of a growing web of movie actor collaboration network. We analyzed the data for corresponding networks and constructed vertex degree distribution. We calculated dependences of vertex nearest neighbor's average degree and of the clustering coefficient on the vertex degree. By comparison of theoretical results with results of data analysis we estimated model parameters. It is shown that 91% of new links connect existing nodes, and only 9% of links come from new actors joining the movie industry.