

ОЦЕНКА ПРИМЕНИМОСТИ МОДЕЛЕЙ РАСТУЩЕГО ГРАФА ДЛЯ МОДЕЛИРОВАНИЯ СЕТИ ЛОКАЛИЗАЦИИ ЗНАНИЙ

Т. Б. Мельникова

Севастопольский филиал Российского экономического университета
имени Г. В. Плеханова, Севастополь, Россия

В статье показано практическое применение растущих графов для построения моделей сети локализации знаний в городах разного масштаба. Сопоставлены результаты моделирования на основе трех типов растущих графов: случайного, предпочтительного присоединения, а также смешанного типа, использующего разные сочетания случайности и предпочтительности при формировании ребра. Для последнего типа растущего графа автором выполнено математическое описание модели. Предложена методология оценки адекватности и эффективности модели, включающая не только выявление формы зависимости реальных и теоретических моделей распределения степеней, динамики степени вершины и среднего локального коэффициента кластеризации, но и анализ внутренней структуры сети. В результате аргументирована невозможность описать реальную сеть локализации знаний единой моделью. Наиболее адекватными стали модели растущего случайного графа и смешанного типа. Динамика коэффициента кластеризации лучше всего моделируется графом предпочтительного присоединения, но с условием несоответствия самого уровня показателя. Полученные выводы актуализируют проблематику выработки моделей для построения малых сетей.

Ключевые слова: растущий граф, знание, город, кластеризация, средняя степень, модель.

ASSESSING APPLICABILITY OF RISING GRAPH MODELS FOR MODELING NETWORK OF KNOWLEDGE LOCALIZATION

Tatyana B. Melnikova

Sevastopol Institute (branch) of the Plekhanov Russian
University of Economics, Sevastopol, Russia

The article shows practical application of rising graphs for building models of knowledge localization net in cities of different scale. Results of modeling based on three types of rising graphs were compared: they are casual, preferred joining and mixed type that uses different combinations of casualty and preference in arc shaping. The author carried out mathematic description of the model for the latter type of rising graph. Methodology of assessing adequacy and efficiency of the model was put forward, which covers not only identifying the form of dependence between real and theoretical models of distribution of degrees, dynamics of node degree and average local clusterization factor but also analysis of the inner structure of net. As a result the author substantiated impossibility to describe the real net of knowledge localization by a uniform model. Models of rising casual graph and of mixed type are considered the most adequate. Dynamics of clusterization factor can be modeled mainly by graph of preferable joining but with condition of disparity of the indicator rate. The obtained conclusions make the objective of developing models for small nets more acute.

Keywords: rising graph, knowledge, city, clusterization, average degree, model.

Введение

Модель растущего графа получила широкое распространение после публикации статьи А.-Л. Бараба-

ши и Р. Альберт, в которой они предложили моделировать развитие сети на основе поэтапного роста изначального графа, состоящего из небольшого количества вер-

шин и ребер, за счет новой вершины, которая через механизм предпочтительного присоединения формирует связь со старыми вершинами [8]. Появление модели было во многом предопределено доступностью большого объема реальных данных благодаря развитию Интернета, поэтому данная модель хорошо отображала отдельные закономерности больших сетей, количество вершин которых насчитывало десятки и сотни тысяч. По мере расширения сферы применения первоначальная модель получила множество модификаций [2].

Среди сфер деятельности, которые изучались посредством моделей растущего графа, особое место занимает наука, где исследуются сети сотрудничества ученых, построенные на основе совместных научных статей, патентов, иных форматов научных работ, а также на основе данных о формировании и развитии групп исследователей и сотрудничестве предприятий в области НИОКР [9; 11; 13]. Все это большие сети, представляющие собой международный либо национальный уровень. Вместе с тем оценка применимости модели для небольших сетей отсутствует. Географическими рамками таких небольших сетей могут быть города или городские агломерации.

В целом можно говорить об исследовании сети локализации знаний на уровне города. В данном случае под локализацией понимается совокупность процессов передачи и накопления знаний внутри какого-то территориального образования. В зависимости от масштаба города такие процессы могут проходить по-разному, что может быть результатом проявления динамических агломерационных эффектов, однако эмпирически проверить влияние таких эффектов достаточно сложно даже для больших городов [5].

Важной для понимания является оценка достаточности накопленного знания у трудовых ресурсов города для ускоренного социально-экономического развития, так как в некоторых работах отмечается дис-

баланс между демографической и производственной пространственной значимостью крупных и средних городов [1].

В целях моделирования локализации знаний на уровне города необходимо построить базовый граф взаимоотношений между агентами, который бы отражал сложившуюся структуру и динамику взаимодействия в процессе передачи знаний. Сеть знаний будем строить на основе соавторства по патентам.

В рамках данной статьи предлагается, во-первых, выявить реальный процесс накопления и передачи знаний между агентами городов разного масштаба, во-вторых, оценить возможность применения растущих графов для моделирования сети локализации знаний на уровне города.

Построение сети локализации знаний городов разного масштаба

Сеть локализации знаний строится по четырем типам городов: большой, средний, крупный, крупнейший¹.

В предыдущих исследованиях нами было предложено сопоставлять сеть знаний городов в рамках единого класса международной патентной классификации (МПК), что позволяет снизить влияние различий в темпах развития между технологиями [5; 6]. В результате для настоящей статьи выбран класс МПК F04D*. Большой, крупный и крупнейший города представлены Приволжским федеральным округом, средний город – Сибирским федеральным округом.

На основе данных Федерального института промышленной собственности произведена выборка по всем патентам на изобретение и полезную модель для выбранных городов. Вершины представлены авторами, ребра – участием в патенте.

Ввиду того, что в реальных сетях регистрация патентов происходила не на ежегодной основе, выполняется расчет коли-

¹ Большой город – от 50 тыс. до 100 тыс. жителей, средний – от 100 тыс. до 250 тыс. жителей, крупный – от 250 тыс. до 1 млн жителей, крупнейший – свыше 1 млн жителей.

чества вершин и ребер в год посредством деления их общего количества на количество лет в периоде выборки патентов (табл. 1).

Т а б л и ц а 1

Параметры сети локализации знаний городов*

Масштаб города	Период выборки патентов	Количество патентов	Количество авторов	Количество вершин в год	Количество ребер в год
Средний	2013–2021	7	23	2,6	16,1
Большой	2006–2023	27	81	4,5	12,3
Крупный	2004–2023	21	20	1,0	1,6
Крупнейший	2005–2023	115	208	10,9	35,5

* Источник: URL: <https://new.fips.ru/iiss/search.xhtml>

Примечание: в выборку входят все патенты на изобретения и полезные модели, которые находятся в одном из трех состояний: действует; может прекратить свое действие; прекратил действие, но может быть восстановлен.

Типы моделей растущих графов

Сравниваются три типа моделей растущего графа: растущий случайный граф, граф Барабаши – Альберт (предпочтительного присоединения) и растущий граф со совмещенным механизмом присоединения. Общими для моделей графов являются изначальные характеристики. На момент времени t_0 мы имеем изначальный набор вершин n_0 , соединенных между собой ребрами (количество ребер меньше количества вершин). На этапе t_1 в сети появляется одна вершина с m ребрами. Далее мы вводим различия:

1. Для растущего случайного графа (модель А): каждое ребро случайным образом выбирает уже имеющуюся вершину. Соответственно, вероятность образования ребра у имеющихся в сети вершин на этапе t равна

$$p = \frac{m}{n_0 + t - 1}.$$

2. Для графа Барабаши – Альберт (модель Б): каждое ребро выбирает имеющуюся вершину в зависимости от позиции ее степени (k_i) относительно других вершин в сети (k_j). Вероятность образования ребра у имеющихся в сети вершин на этапе t равна

$$p_{k_i} = \frac{k_i}{\sum_j k_j}.$$

3. Для растущего графа со совмещенным механизмом присоединения (модель В): на каждом этапе из общего количества

m ребер a ребер присоединяются на основе механизма предпочтительного присоединения, b – на основе случайного механизма (m есть сумма a и b). Соответственно, вероятность образования ребра у имеющихся в сети вершин на этапе t равна

$$p = \frac{b}{n_0 + t - 1} + a \frac{k_i}{\sum_{j=1}^N k_j}. \quad (1)$$

Проверка адекватности и эффективности модели

Адекватность построенной модели непосредственно связана с объектом моделирования. Она должна строиться на максимальной согласованности с исходной информацией, выражаться через ее способность отражать схожую структуру взаимодействия между агентами и отображать свойства, которые являются существенными для реальной сети локализации знаний города. Эффективность модели связана со способностью модели интерпретировать результаты и определяет возможности для получения новой информации об объекте моделирования [3; 4]. При проверке модели графа на ее соответствие реальной структуре взаимоотношения агентов, как правило, опираются на следующие критерии: средняя степень вершины, распределение степеней вершин, коэффициент кластеризации сети в целом, а также динамика степени вершины [8].

Ввиду того что реальные данные могут быть описаны через несколько форм зависимостей со схожим уровнем значимости, введем следующую методологию оценки адекватности модели растущей сети:

1. Для каждого города строятся три типа модели по следующим критериям: распределение степеней вершин, динамика коэффициента кластеризации и динамика вершины – лидера роста (вершина, которая продемонстрировала максимальный темп роста степени).

2. Все теоретические модели строятся на условии появления одной вершины на каждом шаге роста сети. Однако на ежегодной основе сеть всех городов, кроме крупного, увеличивалась более чем на одну вершину. С целью выполнения условия формирования модели индивидуально для каждого города будем проводить корректировку реального времени и количества ребер. Кроме того, для распределения степеней вершин общее количество этапов будет эквивалентно максимальной степени, а для динамики коэффициента кластеризации и динамики вершины – лидера роста – общему количеству лет выборки патентов.

3. Оценка адекватности типа модели опирается на форму зависимости, а также на дополнительный параметр: для распределения степеней вершин – на разницу между наиболее и наименее часто встречающимися степенями в относительном выражении; для динамики коэффициента кластеризации – на разницу между наибольшим и наименьшим значениями коэффициента кластеризации в относительном выражении и его уровнем в конце периода; для динамики вершины – лидера

роста – на разницу между начальным и конечным значениями степени в относительном выражении.

Эффективность модели представляет собой анализ, включающий данные о количестве компонент связности, доле вершин, не меняющих степень, среднем локальном коэффициенте кластеризации на начало и конец периода.

Используя подход, предложенный в работах зарубежных исследователей [8; 10; 12; 14], сформулируем выражения для динамики степени вершины, распределения степеней и среднего локального коэффициента кластеризации для модели В.

Динамика степени вершины. Определим, как будет меняться степень вершины с течением времени. Для этого перейдем от вероятности к непрерывной переменной, стартующей с какого-то изначального значения. Обозначим количество ребер i -й вершины на этапе t через $K_i(t)$ и аппроксимируем данную величину через непрерывную функцию $k_i(t)$, для которой $k_i(t_i) = m$:

$$\frac{dk_i(t)}{dt} = \frac{b}{n_0 + t - 1} + a \frac{k_i(t)}{\sum_{j=1}^N k_j}.$$

При t значительно большем n_0 мы можем выполнить замену $(n_0 + t - 1) \rightarrow t$:

$$\sum_{j=1}^N k_j = 2mt - m = m(2t - 1); \quad m(2t - 1) \rightarrow 2mt,$$

а также составить следующее уравнение для определения динамики степени вершины i :

$$\frac{dk_i(t)}{dt} = \frac{b}{t} + a \frac{k_i(t)}{2mt} = \frac{2mb + ak_i(t)}{2mt}.$$

Решим уравнение

$$k_i(t) = (t^{\frac{a}{2m}} ((am + 2mb) \frac{1}{t_i^{\frac{a}{2m}}} - 2mb) \frac{1}{a} = \frac{m(a + 2b)}{a} (\frac{t}{t_i})^{\frac{a}{2m}} - \frac{2mb}{a}. \quad (2)$$

Распределение степеней вершины. На основе выражения (2) определим вероятность,

что случайно выбранная вершина имеет степень, меньшую k :

$$P[k_i(t)|k] = P\left[\frac{m(a+2b)}{a}\left(\frac{t}{t_i}\right)^{\frac{a}{2m}} - \frac{2mb}{a} \middle| k\right] = P[t_i|t]\left(\frac{ak+2mb}{am+2mb}\right)^{-\frac{2m}{a}}.$$

Разделим полученное в правой части выражение на общее количество вершин N и получим долю вершин, которая удовлетворяет условию (т. е. имеет степень, меньшую k). Принимая во внимание, что $t \rightarrow \infty$ ($N = t$), и используя кумулятивное распределение, получаем

$$P(k) = 1 - \left(\frac{ak+2mb}{am+2mb}\right)^{-\frac{2m}{a}}.$$

Теперь мы можем аппроксимировать данное выражение к доле вершин, имеющих степень k , взяв производную

$$p_k = \frac{dP(k)}{dk} = \left(1 - \left(\frac{ak+2mb}{am+2mb}\right)^{-\frac{2m}{a}}\right)' dk = -\left(\left(\frac{ak+2mb}{am+2mb}\right)^{-\frac{2m}{a}}\right)' dk.$$

Применим правило нахождения производной сложной функции и производной степенной функции:

$$p_k = 2m \frac{\frac{2m+a}{a}}{(a+2b)^{\frac{2m}{a}}} \frac{\frac{2m}{a}}{(ak+2mb)^{\frac{2m+a}{a}}}.$$

Средний локальный коэффициент кластеризации. Коэффициент кластеризации представляет собой вероятность, что два соседа случайно выбранной вершины соединены друг с другом. Мы можем выразить данный подход через локальный коэффициент кластеризации

$$C_s(N) = \frac{2E(S)}{k_s(N)[k_s(N) - 1]}.$$

В числителе отображено актуальное количество связей между соседями вершины s или, иными словами, ожидаемое количество треугольников, в которых участвует вершина s ; в знаменателе – общее количество пар соседей вершины s . Средний локальный коэффициент кластеризации представляет собой следующее выражение:

$$\bar{C}(N) = \frac{\sum_{s=1}^N C_s(N)}{N} = \frac{\int_1^N \frac{2E(S)}{k_s^2(N)} ds}{N}. \quad (3)$$

Основной расчет связан с ожидаемым количеством треугольников для вершины s , т. е. с одновременным наступлением

трех событий: вероятности образования ребра между вершиной s и соседней с ней вершиной i ; вероятности образования ребра между вершиной s и соседней с ней вершиной j ; а также вероятности образования ребра между вершинами i и j .

В отличие от подхода, используемого для расчета модели Б, который принимает во внимание симметричность вероятности образования ребра между двумя вершинами, для растущей случайной сети и, соответственно, модели В необходимо рассмотреть все случаи очередности появления вершин.

Случай 1. Очередность появления вершин $i \rightarrow s \rightarrow j$.

На основе выражения (1) вероятность образования ребра между i и s составляет

$$p(i, s) = \frac{b}{n_0 + s - 1} + a \frac{k_i(s)}{\sum_{j=1}^s k_j(s)}.$$

Вместо t_s , которое обозначает этап появления вершины s , мы используем индекс вершины, т. е. s . Выполним замены: $n_0 + s - 1$ заменим на s :

$$\sum_{j=1}^s k_j(s) = 2ms,$$

$$p(i, s) = \frac{b}{s} + a \frac{m\left[(a+2b)\left(\frac{s}{i}\right)^{\frac{a}{2m}} - 2b\right]}{a} \frac{1}{2ms} = \frac{1}{2} (a+2b) s^{\frac{a-2m}{2m}} i^{-\frac{a}{2m}}.$$

Тогда в сети размера N ожидаемое количество треугольников для вершины s при очередности $i \rightarrow s \rightarrow j$ составит следующее выражение:

$$\begin{aligned} \int_1^s p(i, s) di \int_s^N p(s, j) p(i, j) dj &= \int_1^s \frac{1}{2} (a+2b) s^{\frac{a-2m}{2m}} i^{-\frac{a}{2m}} di \int_s^N \frac{1}{2} (a+2b) j^{\frac{a-2m}{2m}} s^{-\frac{a}{2m}} \frac{1}{2} (a+2b) j^{\frac{a-2m}{2m}} i^{-\frac{a}{2m}} dj = \\ &= \frac{1}{8} (a+2b)^3 \frac{m^2}{(a-m)(m-a)} \left(s^{-\frac{a}{m}} N^{\frac{a-m}{m}} - s^{-1} N^{\frac{a-m}{m}} - s^{-1} + s^{\frac{a-2m}{m}} \right). \end{aligned}$$

Случай 2. Очередность появления вершин $s \rightarrow i \rightarrow j$:

$$\begin{aligned} \int_s^N p(s, i) di \int_i^N p(s, j) p(i, j) dj &= \int_s^N (a+2b) i^{\frac{a-2m}{2m}} s^{-\frac{a}{2m}} di \int_i^N \frac{1}{2} (a+2b) j^{\frac{a-2m}{2m}} s^{-\frac{a}{2m}} \frac{1}{2} (a+2b) j^{\frac{a-2m}{2m}} i^{-\frac{a}{2m}} dj = \\ &= \frac{1}{8} (a+2b)^3 \frac{m}{a-m} \left[s^{-\frac{a}{m}} N^{\frac{a-m}{m}} (\ln N - \ln s - \frac{m}{a-m}) + s^{-1} \frac{m}{a-m} \right]. \end{aligned}$$

Случай 3. Очередность появления вершин $i \rightarrow j \rightarrow s$:

$$\begin{aligned} \int_1^s p(i, s) di \int_i^s p(j, s) p(i, j) dj &= \int_1^s \frac{1}{2} (a+2b) s^{\frac{a-2m}{2m}} i^{-\frac{a}{2m}} di \int_i^s \frac{1}{2} (a+2b) s^{\frac{a-2m}{2m}} j^{-\frac{a}{2m}} \frac{1}{2} (a+2b) j^{\frac{a-2m}{2m}} i^{-\frac{a}{2m}} dj = \\ &= \frac{1}{8} (a+2b)^3 \frac{m}{(m-a)} \left[\frac{m}{(m-a)} s^{-1} - \frac{m}{(m-a)} s^{\frac{a-2m}{m}} - s^{\frac{a-2m}{m}} \ln s \right]. \end{aligned}$$

Ожидаемое количество связей между соседями вершины s :

$$\begin{aligned} E(s) &= \frac{1}{8} (a+2b)^3 \frac{m^2}{(a-m)(m-a)} \left(s^{-\frac{a}{m}} N^{\frac{a-m}{m}} - s^{-1} N^{\frac{a-m}{m}} - s^{-1} + s^{\frac{a-2m}{m}} \right) = \\ &= \frac{1}{8} (a+2b)^3 \frac{m}{a-m} \left[s^{-\frac{a}{m}} N^{\frac{a-m}{m}} (\ln N - \ln s - \frac{m}{a-m}) + s^{-1} \frac{m}{a-m} \right] + \\ &+ \frac{1}{8} (a+2b)^3 \frac{m}{(m-a)} \left[\frac{m}{(m-a)} s^{-1} - \frac{m}{(m-a)} s^{\frac{a-2m}{m}} - s^{\frac{a-2m}{m}} \ln s \right]. \end{aligned}$$

При $N \rightarrow \infty$ ожидаемое количество связей между соседями вершины s будет стремиться к следующему выражению:

$$E(s) = \frac{1}{8} (a+2b)^3 \frac{m^2}{(a-m)^2} s^{-1}. \quad (4)$$

Перед тем как подставить (4) в (3), произведем расчет одного из элементов:

$$k_s(N) [k_s(N) - 1] = k_s^2(N) = \left[\frac{m(a+2b)}{a} \left(\frac{N}{s} \right)^{\frac{a}{2m}} - \frac{2mb}{a} \right]^2 = \frac{m^2(a+2b)^2}{a^2} \left(\frac{N}{s} \right)^{\frac{a}{m}}.$$

В результате

$$\bar{C}(N) = \frac{1}{4} \frac{am(a+2b)}{(a-m)^2} \left(N^{-1} - \frac{1}{N^{\frac{a+m}{m}}} \right).$$

В табл. 2 систематизирована математическая запись моделей. С помощью про-

граммы Gephi произведен расчет средней степени вершины и среднего локального коэффициента кластеризации для реальной сети.

Т а б л и ц а 2

Исходные формулы для моделирования

	Модель А	Модель Б	Модель В
Распределение степеней вершин	$p_k = \frac{e}{m} * e^{-\frac{k}{m}}$	$p_k = 2m^2 k^{-3}$	$p_k = 2m^{\frac{2m+a}{a}} (a+2b)^{\frac{2m}{a}} (ak+2mb)^{-\frac{2m+a}{a}}$
Коэффициент кластеризации	$\bar{C}(N) = 6m[\frac{1}{N} - \frac{1}{N(\ln N + 1)}]$	$\bar{C}(N) = \frac{m}{4} \frac{(\ln N)^2}{N}$	$\bar{C}(N) = \frac{1}{4} \frac{am(a+2b)}{(a-m)^2} (N^{-1} - \frac{1}{N^{\frac{a+m}{m}}})$
Динамика степени вершины	$k_i(t) = m \ln(e^{\frac{n_0+t-1}{n_0+t_i-1}})$	$k_i(t) = m(\frac{t}{t_i})^{1/2}$	$k_i(t) = \frac{m(a+2b)}{a} (\frac{t}{t_i})^{\frac{a}{2m}} - \frac{2mb}{a}$

Построение теоретических моделей и оценка траектории графиков распределения степеней вершин, динамики коэффициента кластеризации и степени автора – лидера роста выполнены в программе Excel.

Результаты исследования

Анализ внутренней структуры реальной сети локализации знаний представлен в табл. 3.

Т а б л и ц а 3

Анализ сети локализации знаний городов

Масштаб города	Количество компонент связности	Средний локальный коэффициент кластеризации на начало/конец периода	Доля вершин, не меняющих степень, %	Максимальное количество патентов у автора
Средний	1	1,0/0,932	58	19
Большой	15	1,0/0,845	80	17
Крупный	4	1,0/0,788	65	6
Крупнейший	23	0,916/0,802	73	40

Средний город (табл. 4). Описание реальной сети. Данный город имеет тип закрытого административно-территориального образования, что может оказать влияние на более высокий уровень кластеризации. Средний локальный коэффициент кластеризации равен 0,932 (табл. 3). Снижение коэффициента кластеризации происходит медленно и описывается степенной, экспоненциальной или логарифмической зависимостями с высоким уровнем значимости (79%). В среднем в год в сеть добавляется 2,6 автора. Динамика степени автора – лидера роста подчиняется экспоненциальному или линейному закону, но с уровнем значимости 30%: за 9 лет она увеличилась в 1,3 раза. Распределение степеней вершины слабо подчиняется какому-либо закону и носит случайный характер. Сеть представляет собой единую компоненту связности.

Чуть более половины вершин (58%) не меняли свою степень.

Построение теоретических моделей и оценка адекватности. Ввиду того что в год сеть увеличивается на 2,6 вершины, тогда шаг роста сети равен 6 месяцам. Исходя из табл. 1 получаем, что каждые полгода появляется в среднем 1,3 вершины (~ 1 вершина) и 8 ребер (соответственно, $m = 8$). В рамках выбора соотношения между количеством случайных ребер (тип b) и тех, которые присоединяются на основе предпочтительного присоединения (тип a), сравнивались 7 вариантов. Вариант $a = 1$, $b = 7$ продемонстрировал превышение наиболее часто встречающейся степени вершины над наименее часто встречающейся в 12,3 раза, что приближено к реальному значению (9 раз). Наиболее часто встречающаяся степень превышает наиме-

нее часто встречающуюся для модели А в 9,5 раза, модели Б – в 6,8 тыс. раз. Соответственно, модель А является наиболее приближенной для распределения степеней в сети знаний среднего города. Однако общая оценка адекватности данной модели

отрицательная ввиду того, что входит в противоречие по распределению степеней: в реальной сети количество вершин с большими степенями не снижается, а возрастает.

Т а б л и ц а 4

Результаты моделирования для среднего города

	Распределение степеней вершин	Динамика коэффициента кластеризации	Динамика степени вершины – лидера роста	Максимальный рост степени, раз
Реальная сеть	$y = 0,11x + 0,16$ (*)	$y = 2E + 52x^{-15,83}$ $y = 7E + 06e^{-0,008x}$ $y = -15,2\ln(x) + 116,6$ (****)	$y = 2E - 13e^{0,01x}$ $y = 0,27x - 522,42$ (**)	1,3
Модель А	$y = 0,34e^{-0,125x}$ (****) $[y = -0,01x + 0,26$ (****)]	$y = 19,39x^{-0,77}$ (****)	$y = 7,37x^{0,27}$ $y = 0,48x + 8,33$ (****)	2,1
Модель Б	$y = 128x^{-3}$ (****)	$y = 0,66x^{0,17}$ (***)	$y = 8x^{0,5}$ (****)	4,2
Модель В	$y = 0,41e^{-0,14x}$ $y = -0,13\ln(x) + 0,39$ (****)	$y = -0,78\ln(x) + 2,78$ (****)	$y = 8,29\ln(x) + 7,55$ (****)	3,9

Примечание. В табл. 4–7 для реальной сети модели построены на ежегодной основе. В год отсутствия изменений берется значение предыдущего этапа. Полиномиальная линия тренда при выборе не используется. Для всех сетей: (*) – коэффициент детерминации ниже 10%; (**) – коэффициент детерминации от 10 до 30%; (***) – коэффициент детерминации от 31 до 70%; (****) – коэффициент детерминации от 71 до 100%. Для моделей А, Б и В указаны уравнения тренда, имеющие наибольшее значение уровня значимости; в квадратных скобках указаны уравнения тренда для выбранной модели, которые соответствуют уравнению тренда реальной сети, но имеют меньшее значение уровня значимости.

Выбор модели для описания динамики коэффициента кластеризации основывается не только на сопоставлении теоретической формы зависимости с реальной, соотношения самого высокого и низкого значений коэффициента, но и на оценке самого уровня коэффициента, который должен быть выше 0,9. В рамках модели В наиболее приближенной можно считать модель, в которой $a = 5$, $b = 3$, динамика кластеризации которой имеет форму логарифмической зависимости и за 18 периодов значение коэффициента кластеризации снизилось на 74%.

При дальнейшем сравнении с моделями А и Б модель Б оказалась наиболее близкой и обладает наименьшим падением значения коэффициента за 18 периодов (на 14%), степенной линией тренда и наиболее высоким уровнем коэффициента кластеризации.

Динамика степени вершины разных типов модели В подчиняется с высокой значимостью логарифмической, степенной и линейной зависимости. Наименьший рост степени наблюдается для варианта $a = 1$, $b = 7$ (3,9 раза). Наиболее подходящей считается модель А с ростом степени в 2,1 раза через 18 периодов и формой зависимости, описываемой в равной степени степенной и линейной, для каждой из которых коэффициент детерминации равен 98%.

Большой город (табл. 5). Описание реальной сети. В городе сеть формировалась с 2006 г., что в два раза больше, чем в среднем городе. По состоянию на 2023 г. в нее входит 81 автор и 27 патентов, что примерно в 4 раза превышает параметры среднего города. В среднем в год сеть увеличивается на 4,5 автора и 12,3 ребра.

Распределение степеней подчиняется в равной степени линейному и логарифмическому законам с уровнем значимости

33%. Максимальное изменение степени составило 3,3 раза за 15 лет, однако динамика реализуется в один шаг и аппроксимируется на протяжении данного периода через степенную или логарифмическую зависимость с уровнем значимости 19%.

Снижение коэффициента кластеризации происходит медленно (на 15% за 15 лет). Сеть сильно фрагментирована и включает 15 изолированных подмножеств (табл. 3). Динамика степени характерна для 20% вершин.

Таблица 5

Результаты моделирования для большого города

	Распределение степеней вершин	Динамика коэффициента кластеризации	Динамика степени вершины – лидера роста	Максимальный рост степени, раз
Реальная сеть	$y = -0,67x + 10,67$ $y = -4,23\ln(x) + 12,99$ (***)	$y = 3E + 66x^{-20,14}$ $y = 5E + 08e^{-0,01x}$ $y = -0,009x + 19,51$ $y = -18,6\ln(x) + 142,41$ (****)	$y = 2E - 200x^{60,74}$ $y = 353,18\ln(x) - 2677,8$ (**)	3,3
Модель А	$y = 0,91e^{-0,33x}$ (****) [$y = -0,03x + 0,41$ (***)]	$y = 9,09x^{-0,88}$ (****)	$y = 3,09x^{0,32}$ (****)	3,9
Модель Б	$y = 18x^{-3}$ (****)	$y = 0,39e^{-0,01x}$ (****) [$y = 0,55x^{-0,21}$ (***)]	$y = 3x^{0,5}$ (****)	8,5
Модель В	$y = 1,03e^{-0,34x}$ (****) [$y = -0,04x + 0,56$ (***)]	$y = 3,04e^{-0,84x}$ (****)	$y = 3,94\ln(x) + 1,28$ (****)	6,2

Построение теоретических моделей и оценка адекватности. Шаг роста сети равен 4 месяцам. В каждый квартал появляется в среднем 1,1 вершины (~ 1 вершина) и 3,1 ребра (соответственно, $m = 3$). Выбор модели В для распределения степеней вершин состоит из двух сочетаний: $a = 1$, $b = 2$ и $a = 2$, $b = 1$. Первый вариант тяготеет к экспоненциальной зависимости и меньшей разнице между наиболее и наименее часто встречающимися степенями. Дальнейшее сравнение трех моделей выявило следующие особенности. Во всех трех моделях наиболее часто встречающиеся степени вершин превышают наименее часто встречающиеся на порядок выше, чем в реальной сети. Однако если мы сопоставим разницу между первым и вторым значениями наиболее часто встречающихся степеней вершин, то для моделей А и В (1,4 и 1,7 соответственно) они приближены к реальному значению (1,5). Обе модели аппроксимируются линейной зависимостью со схожим уровнем значимости, поэтому будем их считать адекватными. Динамика коэффициента кластеризации

лучше всего описывается посредством модели Б, так как она поддерживает сравнительно высокий уровень коэффициента кластеризации с наименьшим его сокращением за 72 периода. Выбор происходил между моделями А и Б ввиду того, что модель В проигрывает в части значения коэффициента кластеризации. В модели Б коэффициент кластеризации сократился за 72 периода на 53%, в модели А – на 94% при реальном сокращении на 15%. В большом городе динамика степени лучше всего описывается моделью А, которая дает рост степени в 3,9 раза (в реальном примере – в 3,3 раза).

Крупный город (табл. 6). Описание реальной сети. Находясь под влиянием большого количества сфер функционирования экономики, город обладает небольшим количеством патентов и авторов в узкой научной сфере. Максимальная степень также самая низкая. Распределение степеней вершин достаточно хаотично и слабо подчиняется только экспоненциальному закону распределения. Динамика коэффициента кластеризации в равной степени мо-

жет быть описана четырьмя законами распределения. Степень автора – лидера роста за 20 лет увеличилась в 2,5 раза. Сеть вклю-

чает 4 изолированных подмножества (табл. 3). Только 35% вершин изменили свою степень.

Таблица 6

Результаты моделирования для крупного города

	Распределение степеней вершин	Динамика коэффициента кластеризации	Динамика степени вершины – лидера роста	Максимальный рост степени, раз
Реальная сеть	$y = 3,98e^{-0,13x}$ (**)	$y = -0,01x + 22,33$ $y = -21,45\ln(x) + 164,08$ $y = 5E + 10e^{-0,01x}$ $y = 3E + 81x^{-24,69}$ (***)	$y = 5E - 42e^{0,05x}$ (****)	2,5
Модель А	$y = 1,36e^{-0,5x}$ (****)	$y = 4,94x^{-0,78}$ (****)	$y = 1,35\ln(x) + 1,89$ (****) [$y = 2,96e^{0,04x}$ (****)]	3,0
Модель Б	$y = 8x^{-3}$ (****)	$y = 0,17x^{0,14}$ (**)	$y = 2x^{0,5}$ (****)	4,5
Модель В	$y = 2,78e^{-0,69x}$ (****)	$y = -0,08\ln(x) + 0,28$ (****)	$y = 2,25x^{0,46}$ $y = 2,32\ln(x) + 1,47$ (****)	4,3

Построение теоретических моделей и оценка адекватности. Шаг роста сети равен одному году, в течение которого появляется 1,0 вершины и 1,6 ребра (в целях сопоставления трех моделей $m = 2$). Модель В имеет только один тип: $a = 1$, $b = 1$.

Согласно уравнениям тренда наиболее близки модели А и В. С одной стороны, в реальной сети разница между наиболее и наименее часто встречающимися степенями составляет 10 раз. Более адекватной считается модель А (12 раз), чем модель В (32 раза). С другой стороны, первое значение наиболее часто встречающейся степени вершин превышает второе в модели В в 2,5 раза, в модели А – в 1,6 раза (в реальной сети – в 3,3 раза), соответственно, модель В становится ближе.

Модель Б наиболее приближенно описывает динамику коэффициента кластеризации, обеспечивая сокращение его значения на 15% (сравнивая с реальным снижением на 21%), однако сам уровень кластеризации очень низкий (0,22).

В модели А за 20 периодов степень вершины – лидера роста выросла в 3 раза, что с учетом уравнения динамики делает эту модель наиболее адекватной.

Крупнейший город (табл. 7). Описание реальной сети. В городе с численностью населения свыше 1 млн человек в несколько раз больше количество патентов, авторов и степень автора – лидера роста. Распределение степеней вершин тяготеет к логарифмической зависимости, а динамика коэффициента кластеризации – практически в равной степени ко всем зависимостям, но наибольшее значение уровня значимости наблюдается для логарифмической и линейной зависимостей. Динамика степени вершины – лидера роста за 19 лет выросла в 5 раз и может быть описана с высоким уровнем значимости экспоненциальной зависимостью.

Отличительной особенностью является поведение среднего локального коэффициента кластеризации, который отличается от единицы на начальном этапе. Ежегодно сеть развивается за счет в среднем 10,9 новых авторов и 35,5 ребер. 73% вершин не меняли свою степень за 19 лет.

Так же, как и в большом городе, сеть локализации знаний крупнейшего города содержит много компонент связности (табл. 3).

Т а б л и ц а 7

Результаты моделирования для крупнейшего города

	Распределение степеней вершин	Динамика коэффициента кластеризации	Динамика степени вершины - лидера роста	Максимальный рост степени, раз
Реальная сеть	$y = -8,64\ln(x) + 29,03$ (***)	$y = -11,2\ln(x) + 86,04$ $y = -0,006x + 12,02$ (***)	$y = 6E - 86e^{-0,09x}$ (****)	5
Модель А	$y = 0,91e^{-0,33x}$ (****)	$y = 10,36x^{-0,92}$ (****)	$y = 2,64\ln(x) + 0,6$ $y = 3,49x^{0,28}$ (****) [$y = 7,99e^{0,004x}$ (***)]	5
Модель Б	$y = 18x^{-3}$ (****)	$y = 0,33e^{-0,01x}$ (****) [$y = -0,08\ln(x) + 0,55$ (****)]	$y = 3x^{0,5}$ (****)	14
Модель В	$y = 0,36e^{-0,23x}$ (****) [$y = -0,19\ln(x) + 0,61$ (***)]	$y = 3,73x^{-0,9}$ (****)	$y = 4,59\ln(x) - 0,67$ $y = 4,74x^{0,31}$ (****)	8

Построение теоретических моделей и оценка адекватности. Шаг роста сети равен 1 месяцу. Каждый месяц появляется в среднем 0,91 вершины (~ 1 вершина) и 2,96 ребра (соответственно, $m = 3$). Выбор модели В состоит из двух сочетаний: $a = 1$, $b = 2$ и $a = 2$, $b = 1$, из которых первое оказалось более близким. В рамках сравнения трех моделей модель В также оказалась более адекватной в части описания соотношения наиболее и наименее часто встречающихся степеней, а также разницы между первым и вторым значениями наиболее часто встречающейся степени вершин. Логика выбора модели Б как лучшего приближения реальной динамики коэффициента кластеризации аналогична предыдущим городам. Рост степени в результате построения модели В составил 8 и 10 раз за 208 периодов, соответственно, для $a = 1$, $b = 2$ и $a = 2$, $b = 1$, что больше максимального роста в реальной сети. Из оставшихся двух моделей модель А продемонстрировала увеличение степени в 5 раз. Вместе с

тем обратим внимание на знак экспоненты, который делает теоретическую модель слабо адекватной для описания реальной сети ввиду того, что в ней скачок динамики сконцентрирован к концу периода, а в теоретической – в начале.

Заключение

Сеть локализации знаний в городах разного масштаба дифференцируется исходя из целого ряда характеристик, включая количество компонент связности, неравномерность появления вершин, различия в темпах роста степени вершин. Эти и другие факторы усложняют унификацию процессов и, соответственно, выбор единой теоретической модели для описания реальных закономерностей. Общий результат показал преобладание случайности в механизме формирования ребра в сети, что, однако же, может свидетельствовать о слабой применимости существующих моделей растущих сетей для малого множества вершин.

Список литературы

1. Власова Н. Ю. Позиции крупных и больших городов в пространственных трансформациях регионов России: кейс Урала // Регион: экономика и социология. – 2022. – № 1 (113). – С. 263–288.
2. Задорожный В. Н., Юдин Е. Б., Юдина М. Н. Методы комплексной калибровки графовых моделей растущих сетей // Системы управления, информационные технологии и математическое моделирование : материалы II Всероссийской научно-практической

конференции с международным участием : в 2 т. – Т. 1. – Омск : Изд-во ОмГТУ, 2020. – С. 26–48.

3. Клейнер Г. Б. Флагман экономико-математического и компьютерного моделирования: 60 лет в строю // Экономика и математические методы. – 2023. – Т. 59. – № 3. – С. 5–20.

4. Клейнер Г. Б. Экономико-математическое моделирование и экономическая теория // Экономика и математические методы. – 2011. – Т. 37. – № 3. – С. 111–127.

5. Мельникова Л. В. Эффективность больших городов: теория и эмпирика // Вопросы экономики. – 2023. – № 3. – С. 83–101.

6. Мельникова Т. Б. Аналитическое и математическое описание особенностей структуры сетей знаний в малых городах // Проблемы развития территории. – 2023. – Т. 27. – № 4. – С. 150–168.

7. Barabási A.-L. Network Science. – Cambridge : Cambridge University Press, 2016.

8. Barabási A.-L., Albert R. Emergence of Scaling in Random Networks // Science. – 1999. – Vol. 286. – P. 509–512.

9. Ducharme L. J., Fujimoto K., Kuo J., Stewart J., Taylor B., Schneider J. Collaboration and Growth in a Large Research Cooperative: A Network Analytic Approach // Evaluation and Program Planning. – 2024. – Vol. 102. – P. 102375.

10. Easley D., Kleinberg J. Networks, Crowds, and Markets: Reasoning about a Highly Connected World. – Cambridge : Cambridge University Press, 2010.

11. König M. D. Technology Cycles in Dynamic R&D Networks // Working Paper. University of Zurich. – 2014. – N 178. – URL: <https://www.econ.uzh.ch/apps/workingpapers/wp/econwp178.pdf>

12. Moriano P., Finke J. Structure of Growing Networks with no Preferential Attachment // Proceedings of the American Control Conference. – 2013. – P. 1088–1093.

13. Piva G. G., Ribeiro L. F., Mata S. A. Networks with Growth and Preferential Attachment: Modelling and Applications // Journal of Complex Networks. – 2021. – Vol. 9. – P. 1–10.

14. Shi D., Zhu S., Liu L. Clustering Coefficients of Growing Networks // Physica A: Statistical Mechanics and its Applications. – 2007. – Vol. 381. – P. 515–524.

References

1. Vlasova N. Yu. Pozitsii krupnykh i bolshikh gorodov v prostranstvennykh transformatsiyakh regionov Rossii: keys Urala [Positions of Large and Big Cities in Spatial Transformations of Regions of Russia: Case of the Urals]. *Region: ekonomika i sotsiologiya* [Region: Economics and Sociology], 2022, No. 1 (113), pp. 263–288. (In Russ.).

2. Zadorozhnyi V. N., Yudin E. B., Yudina M. N. Metody kompleksnoy kalibrovki grafovyykh modeley rastushchikh setey [Methods for Calibration of Models of Growing Networks]. *Sistemy upravleniya, informatsionnye tekhnologii i matematicheskoe modelirovanie: materialy II Vserossiyskoy nauchno-prakticheskoy konferentsii s mezhdunarodnym uchastiem* [Control Systems, Information Technologies and Mathematical Modeling: Proceedings of the II All-Russian Scientific and Practical Conference with International Participation], in 2 Vol. Vol. 1. Омск, Изд-во ОмГТУ, 2020, pp. 26–48. (In Russ.).

3. Kleyner G. B. Flagman ekonomiko-matematicheskogo i kompyuternogo modelirovaniya: 60 let v stroyu [The Flagship of Economic, Mathematical and Computer Modeling: 60 Years in Line]. *Ekonomika i matematicheskie metody* [Economics and Mathematical Methods], 2023, Vol. 59, No. 3, pp. 5–20. (In Russ.).

4. Kleyner G. B. Ekonomiko-matematicheskoe modelirovanie i ekonomicheskaya teoriya [Economic-Mathematical Modeling and Economic Theory]. *Ekonomika i matematicheskie metody* [Economics and Mathematical Methods], 2011, Vol. 37, No. 3, pp. 111–127. (In Russ.).

5. Melnikova L. V. Effektivnost bolshikh gorodov: teoriya i empirika [Efficiency of Large Cities: Theory and Empirics]. *Voprosy ekonomiki*, 2023, No. 3, pp. 83–101. (In Russ.).
6. Melnikova T. B. Analiticheskoe i matematicheskoe opisanie osobennostey struktury setey znaniy v malykh gorodakh [Analytical and Mathematical Description of the Structure Features of Knowledge Networks in Small Cities]. *Problemy razvitiya territorii* [Problems of Territory's Development], 2023, Vol. 27, No. 4, pp. 150–168. (In Russ.).
7. Barabási A.-L. Network Science. Cambridge, Cambridge University Press, 2016.
8. Barabási A.-L., Albert R. Emergence of Scaling in Random Networks. *Science*, 1999, Vol. 286, pp. 509–512.
9. Ducharme L. J., Fujimoto K., Kuo J., Stewart J., Taylor B., Schneider J. Collaboration and Growth in a Large Research Cooperative: A Network Analytic Approach. *Evaluation and Program Planning*, 2024, Vol. 102, p. 102375.
10. Easley D., Kleinberg J. Networks, Crowds, and Markets: Reasoning about a Highly Connected World. Cambridge, Cambridge University Press, 2010.
11. König M. D. Technology Cycles in Dynamic R&D Networks. *Working Paper. University of Zurich*, 2014, No. 178. Available at: <https://www.econ.uzh.ch/apps/workingpapers/wp/econwp178.pdf>
12. Moriano P., Finke J. Structure of Growing Networks with no Preferential Attachment. *Proceedings of the American Control Conference*, 2013, pp. 1088–1093.
13. Piva G. G., Ribeiro L. F., Mata S. A. Networks with Growth and Preferential Attachment: Modelling and Applications. *Journal of Complex Networks*, 2021, Vol. 9, pp. 1–10.
14. Shi D., Zhu S., Liu L. Clustering Coefficients of Growing Networks. *Physica A: Statistical Mechanics and its Applications*, 2007, Vol. 381, pp. 515–524.

Поступила: 17.12.2024

Принята к печати: 16.04.2025

Сведения об авторе

Татьяна Борисовна Мельникова

кандидат экономических наук,
доцент кафедры экономики и управления
Севастопольского филиала
РЭУ им. Г. В. Плеханова.
Адрес: Севастопольский филиал ФГБОУ ВО
«Российский экономический университет
имени Г. В. Плеханова», 299053,
Севастополь, ул. Вакуленчука, д. 29.
E-mail: tmln82@mail.ru
ORCID: 0000-0002-2639-498X

Information about the author

Tatyana B. Melnikova

PhD, Assistant Professor of the Department
for Economics and Management
of the Sevastopol Institute (branch) of the PRUE.
Address: Sevastopol Institute (branch)
of the Plekhanov Russian University
of Economics, 29 Vakulenchuka Str.,
Sevastopol, 299053,
Russian Federation.
E-mail: tmln82@mail.ru
ORCID: 0000-0002-2639-498X