

# АЛГОРИТМ ПРОВЕРКИ ГИПОТЕЗ БИОГРАФИЧЕСКОГО ИССЛЕДОВАНИЯ НА БАЗЕ ГЕОХРОНОЛОГИЧЕСКОГО ТРЕКИНГА

Я. А. Ивакин\*, С. В. Потапычев

Санкт-Петербургский институт информатики и автоматизации Российской академии наук,  
Российская Федерация, Санкт-Петербург, 199178, Российская Федерация

\* Адрес для переписки: [ivakin@oogis.ru](mailto:ivakin@oogis.ru)

## Аннотация

**Предмет исследования.** Информационная технология геохронологического трекинга есть совокупность процессов накопления и интеграции данных о географическом перемещении исторических личностей за установленный период времени с представлением результатов в виде обобщающего графа в ГИС. **Метод.** Гипотезы об устойчивых тенденциях в миграции представимы как подграфы указанного графа. Проверка таких гипотез сведется к поиску и оценке статистической значимости изоморфизма соответствующих графов. Полнофункциональное развитие компьютерной интерпретации методов теории графов на базе геохронологического трекинга способно обеспечить новое качество исторических исследований с использованием современного ГИС-инструментария. **Практическая значимость.** Оно выражается в предоставлении возможности исследователю использовать количественные методы соответствующего логико-аналитического аппарата в своей предметной области. Рассмотрению качественно новых возможностей такого подхода и рационализации соответствующего алгоритмического аппарата посвящена данная статья.

## Ключевые слова

Географические информационные системы, ГИС-технологии для биографических исследований, геохронологический трек и трекинг, изоморфизм графов, рациональный алгоритм, междисциплинарные исследования на базе ГИС.

## Информация о статье

УДК 004.77

Язык статьи – русский.

Поступила в редакцию 22.05.19, принята к печати 02.09.19.

**Ссылка для цитирования:** Ивакин Я. А., Потапычев С. В. Алгоритм проверки гипотез биографического исследования на базе геохронологического трекинга // Информационные технологии и телекоммуникации. 2019. Том 7. № 1. С. 60–74. DOI 10.31854/2307-1303-2019-7-1-60-74.

# ALGORITHM OF BIOGRAPHICAL HYPOTHESES TESTING BASED ON GEOCHRONOLOGICAL TRACKING

Y. Ivakin<sup>1\*</sup>, S. Potapichev<sup>1</sup>

<sup>1</sup> St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg, 199178, Russian Federation

\* Corresponding author: ivakin@oogis.ru

**Abstract—Research subject.** Information technology of the geochronological tracking is an assembly of processes that accumulate and integrate data about geographic relocation of historical figures for a given time interval and represent the results as a generalizing graph in GIS. **Method.** Hypotheses on the stable tendencies in migration could be represented as the above graph's sub-graphs. Such tendencies testing would be reduced to the search and evaluation of the statistical significance for the matching graphs' isomorphism. Full-featured development of computer interpretation of the graph theory methods based on geochronological tracking provides new quality of historical research using modern GIS-tools. **Practical relevance.** Namely, researcher can use the quantitative methods of the corresponding logical-analytical apparatus. The proposed paper deals with a consideration of qualitatively new possibilities of such an approach and the corresponding algorithmic apparatus.

**Keywords—**Geographic information systems, GIS-technologies for biographical research, geochronological track and tracking, graphs' isomorphism, optimal algorithm, GIS based interdisciplinary research.

## Article info

Article in Russian.

Received 10.06.19, accepted 02.09.19.

**For citation:** Ivakin Y., Potapichev S.: Algorithm of Biographical Hypotheses Testing Based on Geochronological Tracking // Telecom IT. 2019. Vol. 7. Iss. 1. pp. 60–74 (in Russian). DOI 10.31854/2307-1303-2019-7-1-60-74

## Введение

Процесс геохронологического трекинга представляет собой совокупность способов сбора первичной биографической, хронологической информации и последовательности приемов обобщения геохронологических треков контролируемых личностей (объектов, артефактов или их совокупностей) на электронной карте (в ГИС). Соответственно геохронологический трек есть интеграция хронологических и географических данных в виде графа, соединяющего географические точки нахождения указанных исторических сущностей (Вершины трека имеют строгое историко-географическое определение, дуги носят характер условно-логической связи). В статьях [1, 2] дано комплексное описание узкоспе-

циализированной программной технологии геохронологического трекинга, а в работе [3] показаны возможности применения аналитического аппарата теории графов и статистических исследований на базе геохронологического трекинга.

Выполнение разработки апробационных примеров построения геохронологических треков по данным из [4, 5] для различных групп исторических личностей позволило перейти к выводу, что финишная версия графа для достаточно представительной выборки исторических существей (личностей, объектов и пр.), как правило, имеет высокосложную и полно- / высокосвязную структуру. Такая структура может быть строго упорядочена. Этот тезис иллюстрирует пример трека, показанный на рис. 1, который построен по результатам геохронологического трекинга данных послужных списков офицеров Главного штаба Военного министерства Российской Армии в период с 1870 по 1910 гг. Именно на основе такого итогового графа геохронологического трекинга становится возможным исследование различных миграционных процессов, выявления некоторых частных исторических закономерностей в перемещении социальных групп, статистически подтвержденная проверка исследовательских гипотез о характере миграций. Существо, концептуальная модель и методологический аппарат таких исследований детально описаны в статье [6].

Концептуальная идея проверки исследовательских гипотез заключается в следующем: итоговый граф геохронотрекинга представляется как граф-базис в структуре которого выявляется подграф изоморфный заданному, т. е. устанавливается наличие взаимно однозначного отображения одного графа на подграф другого, при котором сохраняется отношение инцидентности [7]. Граф, на изоморфность к которому в составе базового графа геохронологического трекинга определяется подграф, топологически описывает ту или иную определенную гипотезу исследования об устойчивой особенности в перемещениях исторических личностей, объектов или других существей в географическом пространстве. Далее определяется степень устойчивости в признании гипотезы исследования о выявляемой особенности в перемещениях с использованием статистического аппарата доверительной вероятности и доверительных интервалов.

Программно-алгоритмическая реализация проверки исследовательских гипотез на базе геохронологического трекинга представляет собой сложную и итеративную вычислительную процедуру. Её практическое воплощение может иметь экспоненциальную сложность и приводить к трансвычислительному характеру решения задачи при определенных входных данных и граничных условиях. Именно этот факт диктует необходимость обоснования и разработки рационального или оптимального алгоритма проверки исследовательских гипотез на базе геохронологического трекинга, т. е. такой локализации вычислительного алгоритма решения задачи определения всех изоморфных вложений в граф геохронотрека, которая позволяет за конечное число подстановок (итераций) определить все комбинации вложений изоморфных заданному графу-гипотезе и не сделать решение трансвычислительным.

Синтез указанного обобщенного алгоритма поиска в граф-базисе подграфа-вложения изоморфного заданному, применительно к специфике входных данных и граничных условий задачи проверки исследовательских гипотез на базе геохронологического трекинга в ГИС, а также обоснование его рационального характера есть предмет рассмотрения данной статьи.

## Конкретизация графовой модели проверки исследовательских гипотез на базе геохронотрекинга

Итогом реализации трекинга, как специализированного программного процесса в ГИС, является географически привязанный граф, обобщающий геохронологические треки отдельных существ или артефактов, информация о миграциях которых занесена в соответствующую базу данных [1]. Именно этот граф является базисом проверки исследовательских гипотез, т. е. в его составе выявляются подграфы изоморфные графам-гипотезам.

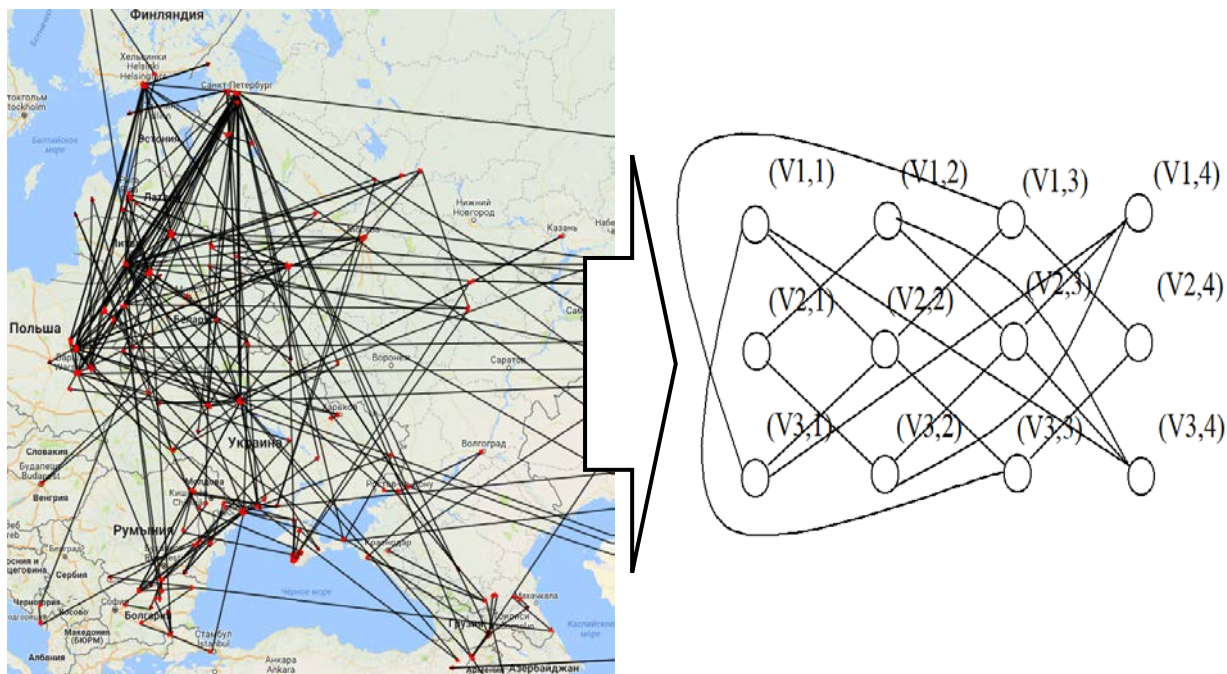


Рис. 1. Рассмотрение геохронологического трека в ГИС в качестве упорядоченного графа

Представленный на рис. 1 пример обобщающего геохронотрека дает возможность понять всю сложность и комбинаторную вариабельность решения частной задачи рационального выделения подграфа изоморфного заданному. Существо указанного выделения заключается в следующем: два графа  $G_1 = (X, U)$  и  $G_2 = (X', U')$  изоморфны, если между парами множеств их вершин, рёбер и дуг существуют взаимно однозначные соответствия, сохраняющие смежность и ориентацию для дуг [3]. Значение попарно изоморфных графов с заданным значением вершин и заданным значением рёбер конечно. Изоморфное отображение  $\varphi$  графа  $G_1$  на граф  $G_2$  задаётся перестановкой

$$\varphi = \begin{pmatrix} x_1 & x_2 & \dots & x_i & \dots & x_p \\ \varphi(x_1) & \varphi(x_2) & \dots & \varphi(x_i) & \dots & \varphi(x_p) \end{pmatrix},$$
 называемой изоморфной, т. е. при распо-

знавании изоморфизма графов  $G_1 = (X_1, R_1)$  и  $G_2 = (X_2, R_2)$  необходимо определить факт изоморфности указанных графов. В случае установления изоморфизма необходимо указать изоморфную подстановку. В виду того, что в теории графов не определены граничные условия и условия оптимальности решения задачи определения строгого соответствия графов, то правомерно применение целого

ряда математических методов установления изоморфизма в качестве методологического базиса разработки искомого алгоритма для случая проверки гипотез исследования на базе геохронологического трекинга.

Условно все множество математических методов установления изоморфизма графов, они же методы решения задачи распознавания изоморфизма графа заданному, можно классифицировать как это показано на рис. 2 [7]. Очевидно, что доминирующими математическими методами в определении изоморфизма графов являются методы, использующие инварианты. (Инвариант графа это некоторая функция, сопоставляющая графу  $L$  соответствующий элемент  $f(L)$  из множества  $M$  элементами которого выступают числа или системы чисел, векторы, многочлены, матрицы и другие алгебраические структуры. Для изоморфных графов значения этой функции совпадают [8]). В соответствии с разнообразием выбора однотипных фрагментов графа различают три класса инвариантов: локальные, квазиглобальные и глобальные.

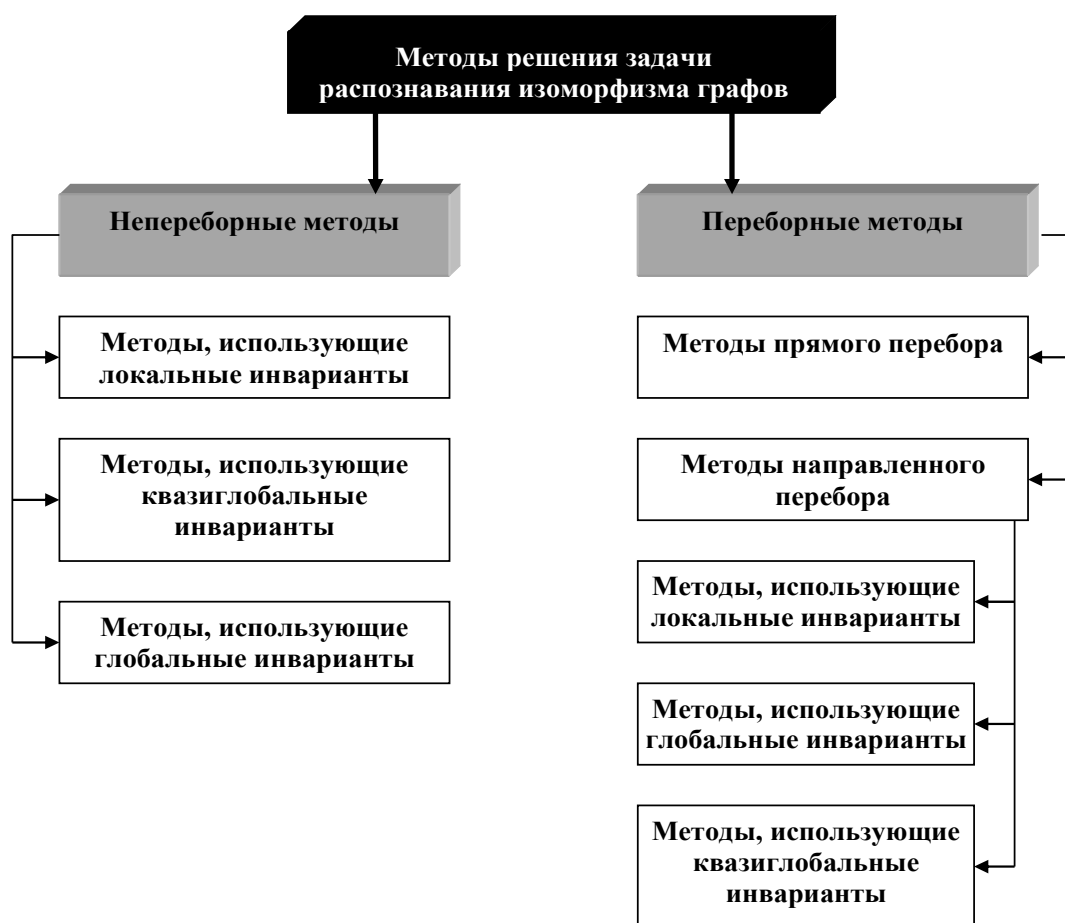


Рис. 2. Классификации математических методов установления изоморфизма графов

При решении задачи поиска изоморфизма графов при различных условиях (размерность графов, их регулярность, однородность и пр.), определяется функция временной сложности самой задачи. Именно эта функция позволяет конкретизировать математический метод решения задачи до прикладного алгоритма, адаптированного к граничным условиям геохронологического трекинга (Упоря-

доченный граф, количество вершин  $n$  от 20 до 100 и пр.). Как правило, это полиномиальный или экспоненциальный алгоритм поиска изоморфизма. Различие между двумя указанными типами алгоритмов особенно заметно при решении задач большой размерности. Данные из таблицы позволяют увидеть причины, по которым полиномиальные алгоритмы более предпочтительны для поиска изоморфизма на геохронотреке по сравнению с экспоненциальными: большинство экспоненциальных алгоритмов – это варианты полного перебора, в то время как полиномиальные алгоритмы возможно построить лишь тогда, когда удаётся строго формализовать предметную суть решаемой задачи. Иными словами, задача строго формализованна если для её решения получен полиномиальный алгоритм [7]. В свою очередь, необходимо показать, что линейное установление изоморфизма графов алгоритмически отличается от более сложной задачи: распознавания изоморфного вложения в составе граф-базиса. Изоморфное вложение или изоморфизм подграфу: Граф  $G_2$  изоморфно вложим в граф  $G_1$ , если в графе  $G_1$  существует подграф, изоморфный графу  $G_2$  [7]. Эта задача отличается от задачи распознавания графов: в частности, чтобы решить задачу изоморфизма подграфа с использованием известных алгоритмов распознавания изоморфизма графов, необходимо реализовать процедуру выявления в графе  $G_1$  подмножества вершин  $X \subset X_1$ , равномошного с множеством вершин  $X_2$  графа  $G_2$ . Данная процедура включает  $k_1$  действий, где  $k_1 = \binom{p_1}{p_2}$ ,  $p_1 = |X_1|$ ,  $p_2 = |x_2|$ . Следовательно,  $k_1$  раз необходимо применить алгоритм распознавания изоморфизма графов как некоторую частную процедуру в составе более общего алгоритма.

Таблица

Временная сложность экспоненциальных алгоритмов решения задачи поиска изоморфизма графов

Функция временной сложности	Размер граф-базиса (Количество вершин $n$ )				
	10	20	30	40	50
$n$	0,00001 с	0,00002 с	0,00003 с	0,00004 с	0,00005 с
$n^2$	0,0001 с	0,0004 с	0,0009 с	0,0016 с	0,0025 с
$n^3$	0,001 с	0,008 с	0,027 с	0,064 с	0,125 с
$n^5$	0,1 с	3,2 с	24,3 с	1,7 мин.	5,2 мин.
$2^n$	0,001 с	1,0 с	17,9 мин.	12,7 дней	35,7 лет
$3^n$	0,059 с	58 с	6,5 лет	3 855 столетий	$2 \times 10^8$ столетий

Таким образом, конкретизация графовой модели проверки гипотез исторического исследования в ГИС, заключается в определении наилучшего математического алгоритма решения задачи изоморфного вложения графов, соответствующего ограничениям и условиям предметной области процедуры геохронотрекинга.

### Рационализация процедуры распознавания изоморфного вложения подграфа-гипотезы в составе граф-базиса

В целях разработки конкретизированного алгоритма распознавания изоморфного вложения подграфа-гипотезы в составе граф-базиса, применительно к задаче проверки исследовательских гипотез в ГИС структуре сводного геохронотрека сопоставлен  $N$ -вершинный граф  $L$ . Пронумеровав его вершины натуральными числами, осуществляется синтез изоморфного к  $L$  графа, вершинами которого служат эти числа. Вариантов такого графа будет  $n!$  – т. е. столько, сколько есть перестановок из  $n$  элементов. Далее задается квадратная матрица:

$$A(L) = \|a_{ij}\|_n$$

с  $n$  строками и  $n$  столбцами, элементы которой определяются как:

$$a_{ij} = \begin{cases} 1, & \text{если } i\text{-я и } j\text{-я вершины в } L \text{ смежны} \\ 0, & \text{если эти вершины не смежны} \end{cases}$$

Очевидно, что  $a_{ii} = 0$  и матрица  $A(L)$  симметрична:  $a_{ij} = a_{ji}$ . Такая матрица будет матрицей смежности графа  $L$  с заданной нумерацией вершин. Т. е., это квадратная матрица  $A = \{a_{ij}\}$ , будет являться матрицей смежности графа  $L$ , если при  $a_{ij} = l$  в графе  $L$  вершины  $x_i$  и  $x_j$  соединены  $l$  рёбрами, при  $a_{ij} = 0$  вершины  $x_i$  и  $x_j$  в  $L$  несмежны.

Степенью  $s(L, x)$  вершины  $x$  графа  $L$  называется число его вершин, смежных с  $x$ , или, что то же, число рёбер, инцидентных этой вершине. При всяком изоморфизме  $\leftrightarrow$  графов  $L$  и  $L'$  соответствующие друг другу вершины должны иметь одинаковую степень, т. е. для любой  $x \in X$  из  $x \leftrightarrow x'$  ( $x' \in X'$ ) должно следовать  $s(L, x) = s(L', x')$ . В самом деле, если для какой-то вершины  $x$  и соответствующей  $x'$  окажется  $s(L, x) > s(L', x')$ , то среди тех  $s(L, x)$  вершин графа  $L'$ , которые отвечают смежным с  $x$  вершинам  $L$ , хотя бы одна не будет смежной с  $x'$ , т. е. соответствие  $\leftrightarrow$  не будет изоморфизмом, но это условие может оказаться приемлемым, в случае если граф  $L'$  изоморфно вложен в граф  $L$ , обратное утверждение ложно.

Так же для ориентированных графов различают полустепени исхода и захода. Полустепень исхода вершины – число инцидентных исходящих дуг,  $S^-(x)$ . Полустепень захода вершины – число инцидентных заходящих дуг,  $S^+(x)$ . Для удобства алгоритмизации полустепени захода вершин записываются в  $n+1$  строке, а полустепени исхода вершин – в  $n+1$  столбце матрицы смежности  $n$  – вершинного ориентированного графа геохронотрека или его подграфов. Перестановочная матрица (матрица подстановок) представляет собой матрицу, в каждой строке и столбце которой находится по одной единице на пересечениях

координат соответствующих друг другу вершин изоморфных графов. При умножении матрицы смежности одного графа слева и справа на перестановочную матрицу можно получить матрицу смежности изоморфного графа [8].

Тогда становится возможным предложить конкретизированный алгоритм распознавания изоморфного вложения подграфа-гипотезы в составе граф-базиса, соответствующий ограничениям и условиям предметной области геохронотрекинга. В соответствии с приведенной выше классификацией методов решения такой задачи математический метод, который реализован в алгоритме, является комбинированным методом направленного перебора. Он объединяет в себе основные преимущества, которые дают методы направленного перебора, использующие локальные, квазиглобальные и глобальные инварианты. В разработанном в ходе исследования алгоритме используются инварианты, как указано ниже:

- число вершин  $n(L)$  ;
- число рёбер  $m(L)$  ;
- вектор степеней  $S(L) = (s_1, s_2, \dots, s_n)$ , который в частности даёт числовые инварианты  $s(L) = \min S(L, x)$  и  $s'(L) = \max S(L, x)$  ;
- полустепень исхода  $S^-(x)$  ;
- полустепень захода  $S^+(x)$  ;
- матрица смежности  $A(L)$  .

При синтезе указанного алгоритма были приняты следующие допущения и ограничения:

- предполагается, что вершины (рёбра) графов имеют одинаковые свойства, т. е. в алгоритме не учитываются весовые коэффициенты вершин (рёбер). Причина в том, что для различных видов графов будут и различные требования к весовым коэффициентам;

- все вершины должны быть пронумерованы;

- рассматриваются только ориентированные графы, т. е. при анализе неориентированных графов необходимо задавать одно ребро как два, связывающих вершины в обоих направлениях;

- не рассматриваются «несвязанные» вершины, т. е. каждое географическое местоположение должно иметь хотя бы одно отношение с другим местоположением (географической точкой);

Методологическим базисом предлагаемого алгоритма решения задачи определения изоморфного вложения графа в граф-базис является тезис о том, что логические схемы  $S_1 = \{A_1, R_1\}$  и  $S_2 = \{A_2, R_2\}$  создают в совокупности систему ограничений, которая воздействует на множество гипотетически возможных вариантов решения и существенно его сокращает. Т. е. производится не перебор вариантов решения, что привело бы к  $N$ -факториальным переборам, а производится наложение системы ограничений по определённому алгоритму на специально созданное поле и на этом поле в результате последовательности действий, направленных на удовлетворение требований этой системы ограничений, формируется уже готовый вариант решения. Этот вариант может выглядеть для графов одинаковой размерности как матрица подстановок. Является допустимым: поле с множеством гипотетически возможных подстановок представить



в виде матрицы размерностью  $n * m$ , для  $n$  и  $m$  – число вершин графов соответственно. Такую матрицу далее корректно называть матрицей возможных подстановок.

Логико-математическое существо матрицы возможных подстановок заключается в том, что на этой матрице отражено всё поле возможных решений текущей задачи изоморфизма графов. Так при решении этой задачи для графов одинаковой размерности путем прямого перебора считается, что каждой вершине  $x_i \in X_1$  графа 1 может быть сопоставлена любая из  $x_j \in X_2$  графа 2. Количество возможных подстановок на матрице размерностью  $n \times n$  будет равным  $n!$ . Эта матрица является удобным средством для фильтрации всех невозможных подстановок. Такая фильтрация имеет два этапа. На этапе номер один производится устранение с поля возможных решений тех вариантов подстановок, которые невозможны принципиально по условию задачи (используя как фильтры глобальные, квазиглобальные и локальные инварианты, а также веса дуг или вершин и др.). На этапе номер два фильтрация вариантов имеет место для выдвигаемых гипотез о той или иной подстановке.

Существо работы предлагаемого алгоритма распознавания изоморфного вложения подграфа-гипотезы в составе граф-базиса, применительно к задаче проверки исследовательских гипотез в ГИС можно эффективно проиллюстрировать на примере решения рассматриваемой задачи для графов, показанных на рис. 3. Матрицы смежности и матрица возможных подстановок для этой пары графов приведены на рис. 4. В матрице возможных подстановок  $C$  в первом столбце перечислены все вершины графа  $B$ , а в первой строке – все вершины графа  $A$ . В столбце  $N + 1$  записывается количество возможных подстановок.

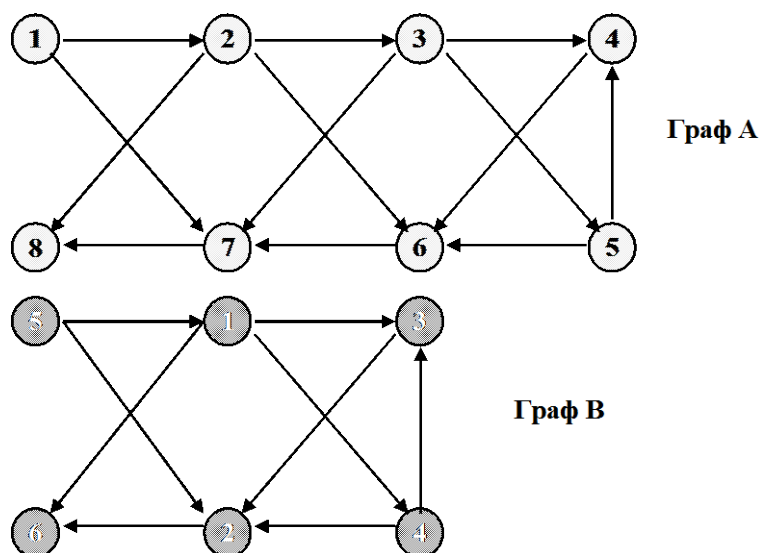


Рис. 3. Пример распознавания изоморфного вложения графа-гипотезы в составе граф-базиса

Первоначальное заполнение матрицы возможных подстановок осуществляется путём анализа полустепеней исхода и захода исходных графов по правилу: вершина может  $B_i$  соответствовать вершине  $A_j$  только в том случае, если полустепени исхода и захода этой вершины больше или равны чем у вершины  $A_j$ . Тем самым производится первая фильтрация вариантов решения. Математически это формализуется и записывается как:

$$C_{ij} = \begin{cases} 1, & \text{если } s(B, b_i) \leq s(A, a_i) \text{ and } s(B, b_j) \leq s(A, a_j) \\ 0, & \text{в остальных случаях} \end{cases}$$

Матрица смежности графа А

	1	2	3	4	5	6	7	8	$S^+(x)$
1	0	1	0	0	0	0	1	0	2
2	0	0	1	0	0	1	0	1	3
3	0	0	0	1	1	0	1	0	3
4	0	0	0	0	0	1	0	0	1
5	0	0	0	1	0	1	0	0	2
6	0	0	0	0	0	0	1	0	1
7	0	0	0	0	0	0	0	1	1
8	0	0	0	0	0	0	0	0	0
$S^+(x)$	0	1	1	2	1	3	3	2	

Матрица смежности графа В

	1	2	3	4	5	6	$S^+(x)$
1	0	0	1	1	0	1	3
2	0	0	0	0	0	1	1
3	0	1	0	0	0	0	1
4	0	1	1	0	0	0	2
5	1	1	0	0	0	0	2
6	0	0	0	0	0	0	0
$S^+(x)$	1	3	2	1	0	2	

Матрица возможных подстановок (матрица С)

	A1	A2	A3	A4	A5	A6	A7	A8	<u>Количество подстановок</u>
B1	0	1	1	0	0	0	0	0	2
B2	0	0	0	0	0	1	1	0	2
B3	0	0	0	1	0	1	1	0	3
B4	0	1	1	0	1	0	0	0	3
B5	1	1	1	0	1	0	0	0	4
B6	0	0	0	1	0	1	1	1	4

Рис. 4. Матрицы смежности графов и возможных подстановок

Дальнейшая корректировка матрицы производится путём наложения на неё следующей системы ограничений:

I. Графы не могут быть изоморфными, если в столбце матрицы «Количество возможных подстановок» имеется хотя бы одна нулевая строка;

II. При однозначном соответствии вершины  $B_i$  вершине  $A_j$  ( $B_k \rightarrow A_j$ ), в столбце  $j$  матрицы возможных подстановок ( $C$ ) не должно быть других символов '1'. Для удовлетворения требований этого ограничения необходимо исключить из матрицы в столбце  $j$  т. н. лишние символы '1'. В связи с тем, что найденное соответствие для какой-либо вершины может оказаться ложным, исключение символов '1' из матрицы необходимо производить таким образом, чтобы оставалась возможность восстановления матрицы на определённом шаге. С этой целью введена переменная *Mirage*, значение которой изменяется после каждого цикла наложения системы ограничений на матрицу возможных подстановок. Из сказанного следуют действия, которые математически можно представить так: если  $C[N + 1, j] = 1$  и  $C[j, l] = '1'$  то элементу  $C[k, l]$ , имеющему значение '1' (для  $k = 1..m; k \neq j$ ), присвоить текущее значение переменной *Mirage*. Если все значения столбца матрицы «количество возможных подстановок» больше 1, то для скорейшего нахождения решения, очевидно, необходимо взять для работы строку с наименьшим значением. А первой, из неиспользованных ранее ячеек соответствующей строки, присваивается т. н. фокус, т. е. назначается активная ячейка (определяются координаты вершин подстановки  $C[FocB, FocA]$ ). Остальные символы '1' необходимо «закрыть» переменной *Mirage*.

III. При  $B_k \rightarrow A_j$ , вершинам  $B_k$  ( $k = 1..m$ ), смежным с  $B_j$ , могут соответствовать только те вершины  $A_l$  ( $l = 1..n$ ), которые смежны с вершиной  $A_j$ . Для удовлетворения этого требования необходимо т. н. лишние символы '1' в матрице возможных подстановок «закрыть» переменной *Mirage*. Математически это можно записать следующим образом:

- a. Если  $C[j, i] = '1'$  и  $B[j, FocB] = '1'$  и  $A[j, FocA] \neq '1'$  то  $C[j, i] = Mirage$ ;
- b. Если  $C[j, i] = '1'$  и  $B[FocB, i] = '1'$  и  $A[FocA, i] \neq '1'$  то  $C[j, i] = Mirage$ ,

где  $A, B$  – исходные матрицы смежности, а  $C$  – матрица возможных подстановок.

IV. Другие частные ограничения, учитывающие специфику распознавания подграфа-гипотезы в составе граф-базиса, применительно к задаче проверки исследовательских гипотез в ГИС.

Приведенный перечень ограничений не является полным и закрытым. В данном случае, ограничения выполняют роль логического фильтра. В зависимости от специфики структуры подграфа, описывающего гипотезу, на структуре граф-базиса могут вводиться т. н. дополнительные фильтры-требования. Тогда текстовое описание работы алгоритма распознавания изоморфного вложения подграфа-гипотезы в составе граф-базиса, применительно к задаче проверки исследовательских гипотез в ГИС, как работы программы решения задачи нахождения изоморфного вложения графов представимо, как показано ниже (для примера графов, показанных на рис. 3):

=> **Начало**

1. Выполнена процедура *Prov\_0\_str*.  $Pr_0 = 0$  переход на *Prov\_1\_str*;
2. *Prov\_1\_str* =>  $Pr_1 = 0$  переход на *Prioritet*;
3. *Prioritet* (с наименьшим числом подстановок две строки (1-я и 2-я). Выбор первой строки и присвоение ей первого приоритета).  $PrTab[1] = 1$ ;  $Mirage = '2'$ ;  $PrEnd = 0$ ; переход на *Mirage1*;
4. *Mirage1*. Определена активная ячейка  $C[1,2]$ , т. е. выдвинута гипотеза, что  $B1 \leftrightarrow A2$ . На основании этой гипотезы получается:  $C[1,3], C[4,2], C[5,2] = '2'$ ;  $PrExit = 0$ ;  $Pr_0 = 0$ . Переход на *Mirage3*.
5. *Mirage3*. В связи с тем, что  $B1$  имеет исходы в  $B3, B4, B6$  ( $B1 \rightarrow B3, B4, B6$ ), а  $A2 \rightarrow A3, A6, A8$ , то, следовательно, и вершинам  $B3, B4, B6$  могут соответствовать только вершины из множества  $A3, A6, A8$ . В таком случае в матрице  $C$  получается:  $C[3,4], C[3,7], C[4,2], C[4,5], C[6,4], C[6,7] = '2'$ ; аналогично для  $B \leftarrow B5$  и для  $A2 \leftarrow A1$ , т.е.  $B5 \leftrightarrow A1$ , а  $C[5,3], C[5,5] = '2'$ ; переход на *Mirage2*;
6. *Mirage2*.  $C[2,6], C[6,6] = '2'$ . Переход на *Prov\_1\_str* => *ZapolnMatrCM* => *ProvEnd*.  $PrExit = 1$ . (Найденная перестановочная матрица оказалась неверной) =>  $Nvar = 0$  => *ExitToHome* (восстановление матрицы  $C$  в прежнем виде) => *Mirage1*;
7. *Mirage1*. Определена активная ячейка  $C[1,3]$ , т. е. выдвинута гипотеза, что  $B1 \leftrightarrow A3$ . На основании этой гипотезы получается:  $C[1,2], C[4,3], C[5,3] = '2'$ ;  $PrExit = 0$ ;  $Pr_0 = 0$ . Переход на *Mirage3*;
8. *Mirage3*. В связи с тем, что  $B1 \rightarrow B3, B4, B6$ , а  $A3 \rightarrow A4, A5, A7$ , то, следовательно, и вершинам  $B3, B4, B6$  могут соответствовать только вершины из множества  $A4, A5, A7$ . В таком случае в матрице  $C$  получается:  $C[3,6], C[4,2], C[6,6], C[6,8] = '2'$ ; аналогично для  $B1 \leftarrow B5$  и для  $A3 \leftarrow A2$ , т. е.  $B5 \leftrightarrow A2$ , а  $C[5,1], C[5,5] = '2'$ ; переход на *Mirage2*;
9. *Mirage2*. Изменений матрицы  $C$  не происходит, переход на *Prioritet*;
10. *Prioritet*. Активной выбирается строка 4.  $Pr = 2$ ; ( $PrTab[4]=2$ );  $Mirage = '3'$ . Переход на *Mirage1*;
11. *Mirage1*. Определена активная ячейка  $C[4,5]$ , т. е. выдвинута гипотеза, что  $B4 \leftrightarrow A5$ . ... Переход на *Mirage3*;
12. *Mirage3*. В связи с тем, что  $B4 \rightarrow B2, B3$ , а  $A5 \rightarrow A4, A6$ , то, следовательно, и вершинам  $B2, B3$  могут соответствовать только вершины из множества  $A4, A6$ . В таком случае в матрице  $C$  получается:  $C[2,7], C[3,7] = '3'$ ; аналогично для  $B4 \leftarrow B1$  и для  $A5 \leftarrow A3$ , т. е.  $B1 \leftrightarrow A3$ , изменений матрицы не происходит; переход на *Mirage2*;
13. *Mirage2*.  $C[6,4] = '3'$ . Переход на *Prov\_1\_str* => *ZapolnMatrCM* => *ProvEnd*.  $PrExit = 1$ . (**решение найдено**) =>  $Nvar = 1$  => *Prioritet* (*попытка поиска автоморфизмов*) =>  $ProvEnd = 1$ ;

=> **Конец.**

При этом матрица возможных подстановок будет в процессе выполнения программы в соответствии с описанным алгоритмом принимать вид, последовательно представленный на рис. 5.

а)

	A1	A2	A3	A4	A5	A6	A7	A8	Количество подстановок
<b>B1</b>		1	2						1
<b>B2</b>						2	1		1
<b>B3</b>				2		1	2		1
<b>B4</b>		2	1		2				1
<b>B5</b>	1	2	2		2				1
<b>B6</b>				2		2	2	1	1

б)

	A1	A2	A3	A4	A5	A6	A7	A8	Количество подстановок
<b>B1</b>		2	1						1
<b>B2</b>						1	1		2
<b>B3</b>				1		2	1		2
<b>B4</b>		2	2		1				1
<b>B5</b>	2	1	2		2				1
<b>B6</b>				1		2	1	2	2

в)

	A1	A2	A3	A4	A5	A6	A7	A8	Количество подстановок
<b>B1</b>		2	1						1
<b>B2</b>						1	3		1
<b>B3</b>				1		2	3		1
<b>B4</b>		2	2		1				1
<b>B5</b>	2	1	2		2				1
<b>B6</b>				3		2	1	2	1

Рис. 5. Динамика изменений вида матрицы возможных подстановок

Из приведённой динамики изменений вида матрицы возможных подстановок видно, что корректировки производились на основе выдвинутых трёх гипотез соответствия пар вершин и заключались, по сути дела, лишь в т. н. подгонке матрицы возможных подстановок под требования системы ограничений для конкретного вида графа-гипотезы.

### Заключение

Использование алгоритмов и вычислительных процедур теории графов применительно к исследовательскому аппарату геохронологического трекинга дает принципиально новые потенциальные возможности для внедрения строгих математических и расчетных методик в сфере ГИС-задач для гуманитарных наук. Также очевидна и перспективность дальнейшей адаптации расчетных методов,

моделей и методик «мягких» вычислений (использование нечетких множеств и нечетких чисел, функций лингвистической переменной и пр.) для ГИС-приложений, применяемых в ходе реализации исторических, этнографических, антропологических и других гуманитарных исследовательских проектов. Данный подход уже сегодня является предметом интереса специалистов в области вычислительной математики, современной алгоритмики и Digital Humanities, что подтверждается такими публикациями как [9, 10, 11, 12, 13, 14, 15, 16].

Дальнейшие направления рационализации алгоритмов проверки гипотез исторического исследования на базе геохронологического трекинга в ГИС связаны с постановкой и решением оптимизационной задачи определения временной сложности указанных алгоритмов, определения строгих граничных условий такой оптимизации и пр. Констатация перспективности описанных направлений позволяет прогнозировать широкое внедрение оптимизированных расчетных алгоритмов в ГИС-инструменты поддержки прикладных исследований.

*Поддержка исследований. Работа выполнена при поддержке РФФИ (проект № 19-07-00006).*

### Литература

1. Ивакин Я. А., Потапычев С. Н. Развитие информационной технологии геохронологического трекинга для исторических исследований в ГИС // Историческая информатика. 2017. № 2 (20). С. 85–94.
2. Ивакин Я. А., Потапычев С. Н. Геохронологический трекинг – специализированный ГИС-инструментарий исторического исследования // Историческая информатика. Информационные технологии и математические методы в исторических исследованиях и образовании. 2016. № 1–2. С. 3–11.
3. Нечепуренко М. И., Попков В. К., Майнагашев С. М., Кауль С. Б., Проскуряков В. А., Кохов В. А., Грызунов А. Б. Алгоритмы и программы решения задач на графах и сетях. Новосибирск: Наука, 1990. 515 с.
4. РГВИА – фонд Ф.400. Главного штаба Военного министерства.
5. РГВИА – фонд Ф.409. Послужные списки, аттестации и наградные листы офицеров русской армии.
6. Ивакин Я. А., Потапычев С. Н., Ивакин В. Я. Проверка гипотез исторического исследования на базе геохронологического трекинга // Историческая информатика. 2018. № 1 (23). С. 86–93.
7. Зыков А. А. Основы теории графов. М.: Вузовская книга, 2004. 664 с.
8. Печенкин В. В., Королев М. С., Дмитров Л. В. Прикладные аспекты использования алгоритмов ранжирования для ориентированных взвешенных графов (на примере графов социальных сетей) // Труды СПИИРАН. 2018. № 6 (61). С. 94–119.
9. ДеМарко Т., Листер Т. Вальсируя с медведями. Управление рисками в проектах по разработке программного обеспечения. М.: ДН, 2005. 196 с.
10. ДеМарко Т. Deadline. Роман об управлении проектами. М.: Манн-Иванов-Фербер, 2016. 352 с.
11. Дюваль П. М., Матиас С., Гловер Э. Непрерывная интеграция. Улучшение качества программного обеспечения и снижение риска. СПб.: Символ, 2016. 240 с.
12. Воротников В. И., Вохмянина А. В. Метод линейаризующей обратной связи в задаче управления по части переменных при неконтролируемых помехах // Труды СПИИРАН. 2018. № 6 (61). С. 61–93.
13. Курейчик В. В., Жиленков М. А. Муравьиный алгоритм для решения оптимизационных задач с явно выраженной целевой функцией // Информатика, вычислительная техника и инженерное образование. 2015. № 2. С. 10–21.
14. Deepak, A., Tobias, F. Average-Case Analysis of Incremental Topological Ordering // Discrete Applied Mathematics. 2010. Vol. 158. No. 4. PP. 240–250.

15. Ammar, A. B. Query Optimization Techniques in Graph Databases // International Journal of Database Management Systems. 2016. Vol. 8. No. 4.
16. Sarma, A. D., Molla, A. R., Pandurangan, G., Upfal, E. Fast Distributed PageRank Computation // Theoretical Computer Science. 2015. Vol. 561. pp. 113–121.

### References

1. Ivakin, Ya. A., Potapychev, S. N. Development of Geochronological Tracking Information Technology for GIS-Based Historical Studies // Historical Informatics. 2017. No. 2 (20). pp. 85–94.
2. Ivakin, Ya. A., Potapychev, S. N. Geochronology Tracking - Specialized GIS-Tool for History Researches // Historical Information Science. Information Technology and Quantitative Methods in Historical Research and Education. 2016. No. 1–2. pp. 3–11.
3. Nechepurenko, M. I., Popkov, V. K., Majngashev, S. M., Kaul', S. B., Proskuryakov, V. A., Kokhov, V. A., Gryzunov, A. B. Algorithms and Programs for Solving Problems on Graphs and Networks. Novosibirsk: Nauka, 1990. 515 p.
4. RGVIA – Fund F.400.
5. RGVIA – Fund F.409.
6. Ivakin, Ya. A., Potapychev, S. N., Ivakin, V. Ya. Testing of Historical Hypotheses on the Basis of Geochronological Tracking // Historical Informatics. 2018. No. 1 (23). pp. 86–93.
7. Zykov, A. A. Fundamentals of Graph Theory. M.: Vuzovskaya kniga, 2004. 664 p.
8. Pechenkin, V. V., Korolev, M. S., Dimitrov, L. V. Applied Aspects of Ranking Algorithms for Oriented Weighted Graphs (on the Example of Social Network Graphs) // SPIIRAS Proceedings. 2018. No. 6 (61). pp. 94–119.
9. DeMarco, T., Lister, T. Waltzing with Bears. Managing Risk on Software Projects. Dorset House, 2003.
10. DeMarco, T. The Deadline. A Novel about Project Management. Dorset House, 1997.
11. Duvall, P. M., Matyas, S., Glover, A. Continuous Integration: Improving Software Quality and Reducing Risk. Addison-Wesley Professional, 2007.
12. Vorotnikov, V. I., Vokhmyanina, A. V. Feedback Linearization Method for Problem of Control of a Part of Variables in Uncontrolled Disturbances // SPIIRAS Proceedings. 2018. No. 6 (61). pp. 61–93.
13. Kureichik, V. V., Zhilenkov, M. A. Ant Colony Algorithms for Solving Optimization Problems with the Explicit Objective Function // Informatika, vychislitel'naya tekhnika i inzhenernoe obrazovanie. 2015. No. 2. pp. 10–21.
14. Deepak, A., Tobias, F. Average-Case Analysis of Incremental Topological Ordering // Discrete Applied Mathematics. 2010. Vol. 158. No. 4. pp. 240–250.
15. Ammar, A. B. Query Optimization Techniques in Graph Databases // International Journal of Database Management Systems. 2016. Vol. 8. No. 4.
16. Sarma, A. D., Molla, A. R., Pandurangan, G., Upfal, E. Fast Distributed PageRank Computation // Theoretical Computer Science. 2015. Vol. 561. pp. 113–121.

**Ивакин  
Ян Альбертович**

– доктор технических наук, профессор,  
ведущий научный сотрудник, СПИИРАН, Санкт-Петербург,  
190178, Российская Федерация, ivakin@oogis.ru

**Потapyчев  
Сергей Николаевич**

– кандидат технических наук, старший научный сотрудник,  
СПИИРАН, Санкт-Петербург, 190178, Российская Федерация,  
potapychev@oogis.ru

**Ivakin Yan**

– Doctor of Engineering Sciences, Full Professor,  
Leading Researcher, SPIIRAS, St. Petersburg, 199178,  
Russian Federation, ivakin@oogis.ru

**Potapichev Sergey**

– Candidate of Engineering Sciences, Senior Research Officer,  
SPIIRAS, St. Petersburg, 199178, Russian Federation,  
potapychev@oogis.ru