

МИНОБРНАУКИ РОССИИ

**Федеральное государственное автономное образовательное учреждение
высшего образования
«Южный федеральный университет»**

**Институт математики, механики и компьютерных наук им.
И.И.Воровича
Кафедра алгебры и дискретной математики**

Мурадян Илья Валерьевич

**ЗАДАЧА ПОИСКА ГРАФА-ПАТТЕРНА
НА ПОМЕЧЕННОМ ГРАФЕ**

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
по направлению 01.03.02 — Прикладная математика и информатика**

**Научный руководитель –
доцент, к.ф.-м.н. Скороходов Владимир Александрович**

Ростов-на-Дону – 2018

Содержание

1	Введение	3
2	Основные понятия	5
3	Задача о поиске паттерна на помеченном графе	7
3.1	Постановка задачи	7
3.2	Алгоритм исключения по локальным условиям	8
3.3	Алгоритм проверки контуров	10
3.4	Общий алгоритм исключения вершин-кандидатов	11
4	Модифицированная задача	11
4.1	Постановка задачи	11
4.2	Измененный алгоритм исключения по локальным условиям	13
4.3	Измененный алгоритм проверки контуров	15
4.4	Рекурсивный алгоритм выделения взвешенного совпадения .	17
5	Список литературы	21

1 Введение

Теория графов - относительно молодой и быстроразвивающийся раздел современной математики. В ней присутствует достаточно много серьёзных теоретических проблем, таких как, например, гипотеза Харари[3] о том, что если граф имеет более трёх рёбер, то его можно однозначно восстановить по подграфам, полученным удалением единственного ребра. Но вместе с этим теория графов успешно применяется для решения прикладных задач, возникающих в теории компьютерных сетей, машинном обучении, при проектировании и эксплуатации транспортных систем, в теории игр и т.д.

Одной из классических задач теории графов является задача об изоморфизме графов. В ней даны два графа G и G' , для которых необходимо построить два биективных отображения δ' и γ таким образом, чтобы они переводили один граф в другой. Понятно, что такие отображения построить можно не всегда. Долгое время для этой задачи не могли найти хорошего алгоритма, и лучшей оставалась оценка $T(n) = \exp(O(\sqrt{n \log n}))$, где n - число вершин, но в 2015 году Ласло Бабай опубликовал статью [5], в которой показывается оценка $T(n) = \exp((\log n)^{O(1)})$, которая очень близка к полиномиальной. Тем не менее, приведённый в этой статье алгоритм всё ещё имеет слишком большую вычислительную сложность для того, чтобы применяться на практике для больших графов. Тем более, задача зачастую стоит немного по-другому: в графе G надо найти частичный подграф [1], изоморфный графу G' , что сразу усложняет любой существующий – переборный или более продвинутый – алгоритм.

В приложениях, однако, часто можно ослабить условие задачи изоморфизма, определенным образом пометив вершины или дуги графов G и G' , то есть сопоставив каждой вершине этих графов какой-то элемент множества L , а каждой дуге – элемент множества M . В этом случае на отображения δ и γ накладываются достаточно жесткие условия: δ может перевести x в x' , а γ – u в u' , только если вершины x и x' и дуги u и u' помечены одинаково. В случае, если множества меток L и M достаточно велики, количество перебираемых для отображений δ и γ вариантов сильно сокращается, что улучшает вычислительную сложность используемых для решения этой задачи алгоритмов.

Задачу изоморфизма помеченных графов также можно поставить более общим и полезным на практике виде: найти такой частичный подграф графа G , который будет изоморфен в обозначенном выше смысле графу G' . Такую задачу называют задачей поиска паттерна на помеченном графе.

В статье [10] приведён алгоритм нахождения граф-паттерна на помеченном графе в случае, если дуги графа не помечены (то есть в формулировке нет множества M), а метки вершин графа G' – уникальны. В подготовленной мной выпускной бакалаврской работе приведено описание этого алгоритма, а также его модификация для случая, когда метки графа G' не уникальны, и допускаются метки на дугах. Такая модификация расширяет применение рассмотренных в статье [10] методов на целый класс прикладных задач. Например, при поиске паттернов на фотографиях часто бывает так, что разыскиваемый паттерн состоит из нескольких одинаковых частей. Таков, например, паттерн лица или солнцезащитных очков. Более того, на таких паттернах зачастую важно учитывать расстояние между их частями, которое отлично моделируется весами на дугах графов-паттернов. В данной работе показано, как учитывать веса на дугах графов-паттернов для нахождения соответствующего данному паттерну частичного подграфа.

Кроме распознавания образов на фотографиях, поиск паттернов на помеченных графах может использоваться и как инструмент социальной инженерии. Так, например, в 2016 году в средствах массовой информации широко освещалась деятельность так называемых групп смерти ([4]), получивших распространение в различных социальных сетях. При наличии полученной административными методами переписки жителей определённого города или района, можно построить граф, вершинами которого будут аккаунты пользователей, а дугами будут связаны те пользователи, которые вели друг с другом переписку. Особым образом следует пометить вершины, соответствующие аккаунтам «целевой аудитории» групп смерти – подростков в возрасте от 13 до 19 лет, а также дуги, соответствующие переписке с различными стоп-словами типа «суицид», «4:20» и пр. После этого следует составить граф-паттерн, выглядящий, например, как граф-звезда (т.е. связный граф, степень всех вершин которого, кроме одной, равна 1 [7]) с 10 листьями-подростками, а дуги графа-паттерна должны быть помечены в обозначенном выше смысле. После того, как паттерн построен, можно применить алгоритм из данной работы и найти все соответствующие ему подграфы, после чего, проводя их анализ, установить координаторов «групп смерти».

Все представленные в данной работе алгоритмы были реализованы на языке Python с использованием библиотеки NetworkX. Оба этих проекта – проекты с открытым исходным кодом и исключительно полной документацией ([9], [11]), которая достаточно активно использовалась в процессе разработки. Для контроля версий и работы за разными машинами также использовалась система контроля версий Git, исходный код которой также

открыт. Несмотря на обилие открытых источников, для более подробного ознакомления с этой системой мною была выбрана книга [8].

2 Основные понятия

Ниже приведены основные понятия и утверждения, необходимые для дальнейшего изложения [1].

Определение 1. Ориентированным графом будем называть двойку $G(V, E)$, где:

- ▷ $V, |V| > 0$ - множество вершин графа;
- ▷ $E \subset V \times V$ - множество дуг графа;

Определение 2. Неориентированным графом будем называть двойку $G(V, E)$, где:

- ▷ $V, |V| > 0$ - множество вершин графа;
- ▷ $E \subset V \times V$ - множество дуг графа;
- ▷ $\forall u, v \in V : (u, v) \in E \Leftrightarrow (v, u) \in E$

Нам также потребуется определение обратных дуг.

Определение 3. Если $e_1 = (u, v) \in E, e_2 = (v, u) \in E$, то дуги e_1 и e_2 называют обратными.

Из определений следует, что неориентированный граф с хотя бы одной дугой содержит по крайней мере одну пару обратных дуг.

Все определения в данном разделе применимы как для ориентированных, так и для неориентированных графов.

Определение 4. Граф $G(V, E)$ будем называть конечным, если множества V и E конечны.

Определение 5. Тройку $G(V, E, \omega)$ будем называть взвешенным графом, если $G'(V, E)$ – граф, а $\omega : E \rightarrow \mathbb{R}$ – весовая функция.

Определение 6. Будем говорить, что дуга $e = (u, v)$ исходит из вершины u и заходит в вершину v . Вершину u будем называть началом, а v – концом дуги.

Определение 7. Будем говорить, что вершина u смежна с вершиной v , если существует дуга $e = (u, v) \in E$. Обозначим множество вершин, смежных с u , через $\Gamma(u)$. Обозначим также $\Gamma^{-1}(u) = \{v | \exists (v, u) \in E\}$. Множество $\text{Inc}(u) = \{(u, v) \in E | v \in V\}$ назовём множеством инцидентных u дуг.

Заметим, что для неориентированных графов для любой вершины u выполнено $\Gamma(u) = \Gamma^{-1}(u)$.

Введём понятия пути и цепи.

Определение 8. Путём называется последовательность дуг $\mu = (e_i = (u_i, u_{i+1}))_{i=0}^{m-1}$. Число m назовём длиной пути и обозначим через $|\mu|$. Будем говорить, что $\mu.s = u_0$ – это начало пути μ , а $\mu.t = u_m$ – это конец пути μ . Путь ненулевой длины μ называется контуром, если его начало совпадает с концом. Простым называется контур, не содержащий одинаковых или обратных дуг. Если граф не содержит простых контуров, он называется бесконтурным.

Также приведём определение цепи.

Определение 9. Цепью на графе G называется последовательность дуг $\lambda = (e_i)_{i=0}^{m-1}$ такая, что для любого $i \in [0; m-2]_{\mathbb{N}}$ дуги e_i и e_{i+1} имеют ровно одну общую вершину (но необязательно начало дуги e_i совпадает с концом e_{i+1}). Число m назовём длиной цепи и обозначим через $|\lambda|$. Будем говорить, что вершина дуги e_0 , не инцидентная дуге e_q – это начало цепи λ , а вершина дуги e_{m-1} , не инцидентная дуге e_{m-2} – это конец цепи λ . Цепь ненулевой длины λ называется циклом, если её начало совпадает с концом.

Необходимо также ввести некоторые определения, специфичные для данной работы.

Определение 10. Расстоянием $d(u, v)$ между вершинами графа u и v называется длина наименьшего пути, началом которого является вершина u , а концом – вершина v , или наоборот. Если ни одного такого пути не существует, то расстояние между вершинами полагается равным бесконечности.

Определение 11. Неориентированным расстоянием $dn(u, v)$ между вершинами графа u и v называется длина наименьшей цепи, началом которой является вершина u , а концом – вершина v . Если ни одной такой цепи не существует, то неориентированное расстояние между вершинами полагается равным бесконечности.

Определение 12. Диаметром $diam(G)$ связного графа G называется максимальное из неориентированных расстояний между его вершинами.

Заметим, что в связном конечном графе неориентированные расстояния между любыми двумя вершинами всегда конечны, поэтому и диаметр такого графа также будет конечен. В данной работе мы будем рассматривать

лишь связные и конечные графы, поэтому все они будут иметь конечный диаметр.

Нам также понадобятся определения, связанные с подграфами и частичными графами.

Определение 13. Граф $G'(V', E')$ называется *частичным графом* графа $G(V, E)$, если $V' = V, E' \subset E$.

Определение 14. Граф $G'(V', E')$ называется *подграфом* графа $G(V, E)$, если $V' \subset V, E' = \{(v, u) \in E | v \in V', u \in V'\}$.

Определение 15. Граф $G'(V', E')$ называется *частичным подграфом* графа $G(V, E)$, если G' является частичным графом некоторого подграфа графа G .

3 Задача о поиске паттерна на помеченном графе

3.1 Постановка задачи

Пусть L – непустое конечное множество (**множество меток**). Пусть $G(V, E), G'(V', E')$ – неориентированные связные графы. Для удобства дальнейшего изложения граф G будем называть *архивным графом*, а граф G' – *графом-паттерном*, или *шаблонным графом*.

Введём отображения $l : V \rightarrow L, l' : V' \rightarrow L$, сопоставляющие вершинам архивного и шаблонного графов соответствующие метки. При этом мы требуем, чтобы введённые на шаблонном графе метки были уникальны, т.е. отображение l' – инъективно.

Для формулировки задачи нам потребуется определение совпадения.

Определение 16. Совпадением на графе G будем называть *частичный подграф* $\hat{G}(\hat{V}, \hat{E})$ графа G такой, что:

1. Существует биективное отображение $m_{\hat{G}} : V' \rightarrow \hat{V}$.
2. $\forall v' \in V' : l'(v') = l(m_{\hat{G}}(v'))$
3. $\forall (v'_1, v'_2) \in E' : (m_{\hat{G}}(v'_1), m_{\hat{G}}(v'_2)) \in E$

Пусть существует некоторое совпадение $\hat{G}(\hat{V}, \hat{E})$ на графе G такое, что есть вершина $v \in V : v \in \hat{V}$. Тогда вершину v будем называть *подходящей паттерну G' по совпадению \hat{G}* , или *истинным кандидатом*, иначе – *неподходящей*. В случае, если вершина v подходит паттерну G' по совпадению \hat{G} , вершину $m_{\hat{G}}^{-1}(v) \in V'$ назовём соответствующей данной вершине v . Заметим, что вершина v может подходить паттерну по нескольким совпадениям, но в силу инъективности отображения l' ей может соответствовать лишь одна вершина $v' \in V'$.

Рассмотрим задачу нахождения множества $\tilde{V} \subseteq V$ всех подходящих паттерну G' вершин.

3.2 Алгоритм исключения по локальным условиям

Для решения поставленной задачи будем использовать следующий алгоритм.

Пусть $T = \{\emptyset\} \cup \mathcal{C}_{V'}^1$, – все 0-элементные и 1-элементные подмножества множества вершин графа-паттерна. Построим отображение $f_0 : V \rightarrow T$, заданное следующим:

$$f_0(v) = \{v' \in V' | l(v) = l'(v')\}. \quad (1)$$

Нетрудно убедиться, что в силу инъективности отображения l , введённое отображение f_0 действительно имеет областью значений множество T .

Алгоритм исключения по локальным условиям будет строиться на изменении отображения f_0 , поэтому для удобства нам потребуется ввести операцию над подобными отображениями. Пусть $f_1, f_2 : V \rightarrow 2^{V'}$. Обозначим $f_2 = \text{Exclude}(f_1, v_0(\in V), v'_0(\in V'))$, если выполнено следующее:

1. $\forall v \neq v_0 \in V : f_1(v) = f_2(v)$.
2. $v'_0 \notin f_2(v_0)$.
3. $\{v'_0\} \cup f_2(v_0) = f_1(v_0)$

Ясно, что по отображению f_1 легко построить отображение $\text{Exclude}(f_1, v_0, v'_0)$, просто исключая вершину v'_0 из множества $f_1(v_0)$.

Прежде чем привести алгоритм, введём ещё несколько определений.

Определение 17. Пусть дан граф $G(V, E)$. Если $\tilde{V} \subset V$, то $\Gamma(\tilde{V}) = \bigcup_{\tilde{v} \in \tilde{V}} \Gamma(\tilde{v})$. В частности, $\Gamma(\emptyset) = \emptyset$.

Определение 18. Пусть $G(V, E)$, $G'(V', E')$ – архивный и шаблонный графы соответственно. Пусть задано отображение $f : V \rightarrow T$, где множество T определено выше. Пусть также $v \in V$. Назовём предикатом локальных условий следующий предикат:

$$LCC(f, v) = \forall u' \in \Gamma(f(v)) \exists u \in \Gamma(v) : u' \in f(u). \quad (2)$$

Алгоритм исключения по локальным условиям принимает на вход отображение f_0 , сопоставляющее вершинам графа их вершины-кандидаты в графе-паттерне. На каждой итерации цикла для каждой вершины $v \in V$ выполнено $f_i(v) \subset f_{i-1}$. На каждой итерации i из списка кандидатов исключаются те вершины v , для которых оказывается, что хотя бы для одной из

вершин $u' \in \Gamma(f_i(v))$ не существует такой вершины $u \in \Gamma(v)$, что $f(u) = u'$. Схожая идея используется в различных классических алгоритмах теории графов, например, в алгоритме Беллмана-Форда [6]. Следуя той же схеме доказательства, которая используется, например, при обосновании того факта, что алгоритм Беллмана-Форда находит на бесконтурном графе кратчайшие пути за количество итераций, равное диаметру графа, мы придём к тому, что для бесконтурного графа G' алгоритм *LCCE* отработает за $diam(G')$ итераций, после чего все вершины $v \in V$, для которых $f(v)$ не пусто, будут действительно вершинами некоего графа-совпадения, ведь для каждой вершины $u = f(v)$, каждая начинающаяся в ней цепь (в силу того, что граф неориентирован, пути и цепи суть оно и то же) имеет некую соотносящуюся с этой цепью цепь в графе G , начинающуюся в вершине v . Для бесконтурного графа можно отдельно не вычислять его диаметр, просто прервав алгоритм после итерации i , для которой $F_i = F_{i-1}$.

Псевдокод алгоритма исключения по локальным условиям приведён ниже (алгоритм 1).

Алгоритм 1: Алгоритм исключения по локальным условиям

Вход: графы $G(V, E)$, $G'(V', E')$, отображение $f_0 : V \rightarrow T$, число итераций N

Выход: изменённое отображение f_K

```

1  начало LCCE
2     $K := 0$ 
3     $F_0 := f_0$ 
4    для  $i = 1, 2, \dots, N$  :
5      для  $v \in V$  :
6        если  $\neg LCC(f_i, v)$  тогда
7           $S := K$ 
8          для  $v' \in f_S(v)$  :
9             $f_{K+1} := Exclude(f_K, v, v')$ 
10          $K := K + 1$ 
11      $F_i := f_K$ 
12  вернуть  $f_N$ 
```

Для графов-паттернов с простыми контурами применение только этого алгоритма не даёт требуемого результата. На таких графах существуют бесконечные пути, а потому для них применение приведённого выше алгоритма со сколь угодно большим числом итераций не гарантирует того, что оставшиеся вершины-кандидаты будут в действительности вершинами некоего графа-совпадения. Поэтому для графов с простыми контурами

мы будем применять, помимо алгоритма исключения по локальным условиям, ещё и алгоритм проверки контуров, описанный ниже.

3.3 Алгоритм проверки контуров

Пусть \mathcal{K}_0 – это набор контуров графа G' . В этот набор достаточно включить все простые контуры этого графа. Пусть $\mathcal{C}_0 \in \mathcal{K}_0$ – какой-то из рассматриваемых контуров. Тогда если $\mathcal{C}_0 = ((v'_0, v'_1), \dots, (v'_{r-1}, v'_0))$ и $v'_0 \in f(v_0)$, то алгоритм проверки контуров 2, приведённый ниже, исключит вершину v'_0 из множества $f(v_0)$, если окажется, что нет контура, начинающегося в вершине v_0 и соответствующего (в смысле отображения f) контуру \mathcal{C}_0 .

Алгоритм 2: Алгоритм проверки контуров

Вход: графы $G(V, E)$, $G'(V', E')$, отображение $f_0 : V \rightarrow T$

Выход: изменённое отображение f_N

```

1  начало ССЕ
2   $K := 0$ 
3  для каждого  $\mathcal{C}_0 \in \mathcal{K}_0$  выполнять
4      Пусть  $(v'_0, v'_1)$  – первая дуга контура  $\mathcal{C}_0$ .
5       $\mathcal{A} := \emptyset$ 
6       $\mathcal{A}_0 := \emptyset$ 
7      для всех  $v_0 : f_K(v_0) \neq \emptyset$  выполнять
8           $\mathcal{A}_0 := \mathcal{A}_0 \cup \{v_0\}$ 
9           $\mathcal{A} := \mathcal{A} \cup \{(v_0, v_0, 0)\}$ 
10     для  $s = 1, 2, \dots, |\mathcal{C}_0|$  :
11         Пусть  $(q_0, q_1)$  –  $s$ -я дуга контура  $\mathcal{C}_0$ .
12          $\mathcal{B} := \emptyset$ 
13         для каждого  $(v, v_0, s-1) \in \mathcal{A}$  выполнять
14             для  $v' \in \Gamma(v)$  :
15                 если  $q_1 \in f_K(v')$  тогда
16                      $\mathcal{B} := \mathcal{B} \cup \{(v', v_0, s)\}$ 
17              $\mathcal{A} := \mathcal{B}$ 
18         для каждого  $v_0 \in \mathcal{A}_0$  выполнять
19             если  $(v_0, v_0, |\mathcal{C}_0|) \notin \mathcal{A}$  тогда
20                  $S := K$ 
21                 для  $v' \in f_S(v_0)$  :
22                      $f_{K+1} := \text{Exclude}(f_K, v_0, v')$ 
23                  $K := K + 1$ 
24  вернуть  $f_N$ 

```

Нахождение всех контуров графа и такой их обход – задачи с очень высокой вычислительной сложностью, но на практически встречающихся графах, тем не менее, основная работа будет сделана алгоритмом *LCCE*. Впрочем, можно даже рассматривать не все циклы – можно попытаться выделить подграф совпадения и по неточному списку кандидатов. Здесь много зависит от конкретного вида графов и стоящей перед нами задачи.

Авторы статьи [10] ожидают, что граф-паттерн будет содержать достаточно мало контуров, поэтому приводимая ими оценка сложности $O(|\mathcal{K}_0| n_t(|V| + |E|))$ является удовлетворительной. Здесь n_t определяется следующим образом. Если n_{C_0} – это количество вершин в V , соответствующих контуру C_0 , то $n_t = \max_{C_0 \in \mathcal{K}_0} n_{C_0}$. Т

3.4 Общий алгоритм исключения вершин-кандидатов

Два вышеприведённых алгоритма объединяются в алгоритме 3, который выполняется до тех пор, пока из отображения f ещё удаляются какие-то вершины-кандидаты.

Алгоритм 3: Алгоритм исключения вершин-кандидатов

Вход: графы $G(V, E)$, $G'(V', E')$, отображение $f_0 : V \rightarrow T$

Выход: изменённое отображение f_N

1 **начало** Exclusion

2 $K := 0$

3 **до тех пор, пока** $K = 0$ *или* $f_K \neq f_{K-2}$ **выполнять**

4 $f_{K+1} := LCCE(G, G', f_K)$

5 $f_{K+2} := CCE(G, G', f_{K+1})$

6 $K := K + 2$

7 **вернуть** f_K

Формируя начальное отображение f_0 исходя из совпадений по меткам архивного графа и графа-паттерна, мы передаём его алгоритму Exclusion и получаем на выходе новое отображение f_N такое, что для любой вершины $v \in V$ каждая вершина множества $f_N(v)$ имеет вершину v истинным кандидатом.

4 Модифицированная задача

4.1 Постановка задачи

Пусть, как и прежде, L – непустое конечное множество, называемое **множеством меток**. Пусть $G(V, E)$, $G'(V', E')$ – ориентированные графы.

Кроме того, будем полагать, что граф G' – связный. Для удобства дальнейшего изложения будем называть граф G архивным графом, граф G' – графом-паттерном, или шаблонным графом.

Введём отображения $l : V \rightarrow L$, $l' : V' \rightarrow L$, сопоставляющие вершинам архивного и шаблонного графов соответствующие метки. Никаких дополнительных требований на эти отображения мы уже не накладываем.

Пусть также \mathcal{X} – это непустое (возможно, бесконечное) множество (**множество характеристик**) с заданным на нём бинарным отношением $\rho : \mathcal{X} \times \mathcal{X} \rightarrow \{0, 1\}$

Мы вводим множество характеристик для того, чтобы, например, работать со взвешенными графами. Действительно, если положить

$$\mathcal{X} = \mathbb{R},$$

$$\rho(x, y) = \begin{cases} 1, & |x - y| < \varepsilon \\ 0, & |x - y| \geq \varepsilon \end{cases}, \varepsilon > 0,$$

мы получим отношение ρ , заданное, как отношение «примерного равенства». Его мы и будем использовать для сравнения весов дуг архивного графа и графа-паттерна.

Введём также «помечающие отображения» для дуг графа: $\chi : E \rightarrow \mathcal{X}$, $\chi' : E' \rightarrow \mathcal{X}$.

Несколько изменится и определение совпадения.

Определение 19. *Взвешенным совпадением на графе G будем называть частичный подграф $\hat{G}(\hat{V}, \hat{E})$ графа G такой, что:*

1. *Существует биективное отображение $m_{\hat{G}} : V' \rightarrow \hat{V}$.*
2. $\forall v' \in V' : l'(v') = l(m_{\hat{G}}(v'))$
3. $\forall e' = (v'_1, v'_2) \in E' : \rho(\chi'(e'), \chi((m_{\hat{G}}(v'_1), m_{\hat{G}}(v'_2)))) = 1$
4. $\forall (v'_1, v'_2) \in E' : (m_{\hat{G}}(v'_1), m_{\hat{G}}(v'_2)) \in E$

Пусть существует некоторое взвешенное совпадение $\hat{G}(\hat{V}, \hat{E})$ на графе G такое, что есть вершина $v \in V : v \in \hat{V}$. Тогда вершину v будем называть подходящей паттерну G' по взвешенному совпадению \hat{G} , или истинным кандидатом, иначе – неподходящей. В случае, если вершина v подходит паттерну G' по взвешенному совпадению \hat{G} , вершину $v' = m_{\hat{G}}^{-1}(v) \in V'$ назовём соответствующей вершине v , а вершину v – соответствующей вершине v' .

Рассмотрим задачу нахождения соответствующих вершин для каждой из вершин графа G' . В следующих двух разделах описаны модификации алгоритмов из статьи [10], решающие задачу в такой постановке. В разделе 4.4 описан алгоритм, восстанавливающий взвешенные совпадения по найденным соответствующим вершинам.

4.2 Измененный алгоритм исключения по локальным условиям

Пусть $T = 2^V$ – все подмножества множества вершин графа-паттерна. Построим отображение $f_0 : V' \rightarrow T$, заданное следующим:

$$f_0(v')\{v \in V | l(v) = l'(v')\}. \quad (3)$$

В отличие от приведённого в предыдущей части работы, это отображение сопоставляет каждой вершине графа-паттерна вершину какое-то подмножество вершин архивного графа. Также это отображение является истинно многозначным, т.е. элементами его образа являются, в общем случае, не только 0- и 1-элементные подмножества множества V .

В связи с изменением области значений и определения введённого отображения f_0 следует изменить и операцию *Exclude*. Пусть $f_1, f_2 : V' \rightarrow 2^V$. Обозначим $f_2 = \text{Exclude}(f_1, v_0(\in V), v'_0(\in V'))$, если выполнено следующее:

1. $\forall v' \neq v'_0 \in V' : f_1(v') = f_2(v')$.
2. $v_0 \notin f_2(v'_0)$.
3. $\{v_0\} \cup f_2(v'_0) = f_1(v'_0)$

Ниже приведён изменённый алгоритм исключения по локальным условиям. Он принимает те же самые графы, но входное и выходное отображение имеют другие области определения и значений. Общая идея алгоритма со-

храняется.

Алгоритм 4: Измененный алгоритм исключения по локальным условиям

Вход: графы $G(V, E)$, $G'(V', E')$, отображение $f_0 : V' \rightarrow T$, число итераций N

Выход: изменённое отображение f_K

```
1  начало MLCSE
2   $K := 0$ 
3   $F_0 := f_0$ 
4  для  $i = 1, 2, \dots, N$  :
5      для  $(q_0, q) \in E'$  :
6          для  $v_0 \in f(q_0)$  :
7               $flag := False$ 
8              для  $v \in \Gamma(v_0)$  :
9                  если  $v \in f_K(q)$  тогда
10                     если  $\rho(\chi(v_0, v), \chi'(q_0, q)) = 1$  тогда
11                          $flag := True$ 
12                 если  $flag = False$  тогда
13                      $f_{K+1} := Exclude(f_K, v_0, q_0)$ 
14                      $K := K + 1$ 
15             для  $v \in f(q)$  :
16                  $flag := False$ 
17                 для  $v_0 \in \Gamma^{-1}(v)$  :
18                     если  $v_0 \in f_K(q_0)$  тогда
19                         если  $\rho(\chi(v_0, v), \chi'(q_0, q)) = 1$  тогда
20                              $flag := True$ 
21                 если  $flag = False$  тогда
22                      $f_{K+1} := Exclude(f_K, v, q)$ 
23                      $K := K + 1$ 
24      $F_i := f_K$ 
25  вернуть  $f_N$ 
```

Алгоритм *MLCSE* полностью берёт идею алгоритма, изложенного выше, а потому его корректность прямо вытекает из корректности вышеизложенного алгоритма. Мы лишь добавили пару циклов для того, чтобы обрабатывать многозначность, а также ориентированность графов G и G' .

Сложность приведённого алгоритма, очевидно, выше сложности алгоритма *LCSE*. Оценим её на одной итерации. Поскольку $|\Gamma(v_0)|, |\Gamma^{-1}(v_0)| \leq$

$|V|$, то

$$T = \sum_{(q_0, q) \in E'} |f_i(q_0)| \cdot |V| = \sum_{q_0 \in V'} |\Gamma(q_0)| \cdot |f_i(q_0)| \cdot |V| \leq$$

В худшем случае $f(q_0) = |V|$, имеем:

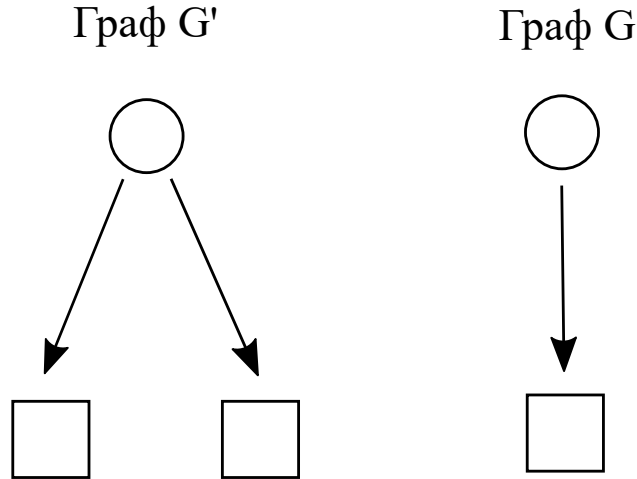
$$\leq |E'| \cdot |V|^2.$$

Таким образом, $T = O(|E'| \cdot |V|^2)$

Худший случай, однако, реализуется, только если задача поиска паттерна вырождается в задачу изоморфизма. Для реальных задач алгоритм *MLCCE* работает достаточно быстро, за несколько итераций уменьшая число кандидатов на несколько порядков.

Заметим, однако, что в некоторых случаях даже на бесконтурных графах алгоритм 4 может не исключать некоторые вершины из списка кандидатов, несмотря на то, что они не являются истинными кандидатами. Простейший пример приведён на рисунке 1. В таких случаях можно, во-первых, применить фильтрацию по степеням вершин, а во-вторых, воспользоваться алгоритмом поиска взвешенного совпадения из последнего раздела данной работы.

Рис. 1: Пример графов G и G' , для которых алгоритм *MLCCE* не исключает неистинных кандидатов



4.3 Измененный алгоритм проверки контуров

Для корректного исключения кандидатов на графах с контурами также был модифицирован алгоритм проверки контуров.

Данный алгоритм проверяет на истинность кандидатов \hat{v}_0 всех вершин v'_0 графа-паттерна, в которых начинается какой-то контур \mathcal{C}_0 . Для этого на каждой итерации цикла по дугам контура \mathcal{C}_0 создаётся множество вершин

\mathcal{B} , в которое добавляются те вершины, которые соединены с вершинами множества \mathcal{A} подходящими по характеристике дугами. Также проверяется, что концы этих дуг графа G являются кандидатами для соответствующих концов рассматриваемой на текущей итерации дуги контура \mathcal{C}_0 . После этого множеству \mathcal{A} присваивается множество \mathcal{B} . Таким образом, после s итераций цикла по дугам множество \mathcal{A} содержит все вершины, на которые оканчиваются пути длины s , начинающиеся в вершине \hat{v}_0 и являющиеся возможными взвешенными соответствиями пути, состоящего из первых s дуг контура \mathcal{C}_0 . Отдельно обрабатывается случай последней дуги контура, конец которой должен совпадать с началом первой дуги.

Псевдокод изменённого алгоритма проверки контуров приведён ниже (алгоритм 5).

Алгоритм 5: Изменённый алгоритм проверки контуров

Вход: графы $G(V, E)$, $G'(V', E')$, отображение $f_0 : V' \rightarrow T$

Выход: изменённое отображение f_N

```

1  начало МССЕ
2   $K := 0$ 
3  для каждого  $\mathcal{C}_0 \in \mathcal{K}_0$  выполнять
4      Пусть  $(v'_0, v'_1)$  – первая дуга контура  $\mathcal{C}_0$ .
5      для  $\hat{v}_0 \in f_K(v'_0)$  :
6           $\mathcal{A} := \{\hat{v}_0\}$ 
7          для  $s = 1, 2, \dots, |\mathcal{C}_0|$  :
8              Пусть  $(q_0, q_1)$  –  $s$ -я дуга контура  $\mathcal{C}_0$ .
9               $\mathcal{B} := \emptyset$ 
10             для каждого  $v_0 \in \mathcal{A}$  выполнять
11                 для каждого  $v_1 \in f_K(q_1)$  выполнять
12                      $flag := False$ 
13                     если  $v_1 \in \Gamma(v_0)$  тогда
14                         если  $s \neq |\mathcal{C}_0| \vee \hat{v}_0 = v_1$  тогда
15                             если  $\rho(\chi(v_0, v_1), \chi'(q_0, q_1)) = 1$  тогда
16                                  $flag := True$ 
17                     если  $flag = True$  тогда
18                          $\mathcal{B} := \mathcal{B} \cup \{v_1\}$ 
19              $\mathcal{A} := \mathcal{B}$ 
20             если  $\hat{v}_0 \notin \mathcal{A}$  тогда
21                  $f_{K+1} := Exclude(f_K, \hat{v}_0, v'_0)$ 
22              $K := K + 1$ 
23  вернуть  $f_N$ 

```

Поскольку для любой вершины $v' \in V'$ $|f(v')| \leq |V|$, а также $|\mathcal{A}| \leq |V|$ и $|\mathcal{C}| \leq E'$, сложность измененного алгоритма можно грубо оценить как $T = O(|\mathcal{K}| \cdot |V|^3 \cdot |E'|)$.

4.4 Рекурсивный алгоритм выделения взвешенного совпадения

Общий алгоритм без каких-либо изменений переносится с уже введённого выше (алгоритм 3). Разница лишь в том, что передаваемое ему отображение формируется так, как указано выше в этом разделе.

Таким образом, построен алгоритм, решающий гораздо более общую задачу. Однако полученное после его работы отображение f_N не может использоваться непосредственно – необходимо ещё выделить взвешенные совпадения на графе G . Для решения этой задачи будем использовать описанный ниже алгоритм.

Для восстановления взвешенных совпадений по полученному отображению f был разработан следующий метод:

1. Преобразовать граф-паттерн G' к особому виду, снабдив его дополнительной информацией, получив таким образом граф G_0 .
2. Положить $i := 0$.
3. Пока граф G_i имеет более одной вершины:
 - (a) Выполнить на графе G_i обход в ширину с окраской.
 - (b) Выделить подграфы графа G_i с вершинами одного цвета.
 - (c) На каждом выделенном подграфе $G_i^j(V_i^j, E_i^j)$ построить все возможные соответствия между его вершинами и вершинами архивного графа.
 - (d) Сформировать новый граф G_{i+1} , взяв в качестве вершин подграфы G_i^j , а в качестве дуг – дуги, вершины которых принадлежат сразу нескольким подграфам.
 - (e) Положить $i := i + 1$.

Рассмотрим подробнее каждый из этих пунктов.

Преобразование графа G' выполняется следующим образом: каждая его вершина v' получает как характеристику список соответствий $\xi(v')$. Каждый список соответствий состоит из кортежей, элементами которых являются пары вида (q_0, v_0) , означающих, что вершине q_0 графа-паттерна соответствует вершина v_0 архивного графа. Таким образом, изначально для каждой вершины $v' \in V'$:

$$\xi(v') = [((v', v_1)), \dots, ((v', v_L))], v_1, \dots, v_L \in f(v_L).$$

Граф G' со списками соответствий ξ будем считать начальным графом для алгоритма выделения взвешенных соответствий и обозначать G_0 .

Опишем алгоритм поиска в ширину с окраской, необходимый для решения задачи разбиения графа G_i на подграфы. Процедура BFS_C , приведённая ниже (алгоритм 6), окрашивает максимум N вершин графа G в порядке обхода в ширину, начиная со стартовой вершины s .

Алгоритм 6: Обход в ширину с окраской

Вход: граф $G(V, E)$, вершина $s \in V$, отображение $c : V \rightarrow \mathbb{Z}_+$, цвет i , лимит вершин N

Выход: изменённое отображение c

```

1 начало  $BFS_C$ 
2    $n := 0$ 
3   Пусть  $Q$  – пустая очередь
4   Добавить  $s$  в  $Q$ 
5   до тех пор, пока очередь  $Q$  не пуста и  $n < N$  выполнять
6     Извлечь вершину  $v$  из очереди  $Q$ .
7      $c(s) := i$ 
8      $n := n + 1$ 
9     для всех  $v_0 \in \Gamma(v)$  выполнять
10      если  $c(v_0) \neq 0$  тогда
11        Добавить вершину  $v_0$  в очередь  $Q$ .
12 вернуть  $c$ 
```

Процедуру 6 используем в общем алгоритме окраски графа 7.

Алгоритм 7: Общий алгоритм окраски графа

Вход: граф $G(V, E)$, лимит вершин N

Выход: окрашивающее отображение c , количество цветов K

```

1 начало  $BFSColoring$ 
2   для  $v \in V$  :
3      $c(v) := 0$ 
4    $i := 0$ 
5   до тех пор, пока  $\exists v : c(v) = 0$  выполнять
6     Пусть вершина  $v$  такова, что  $c(v) = 0$ 
7      $i := i + 1$ 
8      $c := BFS_C(G, v, c, i, N)$ 
9   вернуть  $c, i$ 
```

После окраски по тем вершинам графа G_i , для которых алгоритм 7 определил одинаковые цвета, выделим подграфы G_i^1, \dots, G_i^K .

В каждом выделенном подграфе отыщем все взвешенные совпадения,

для чего воспользуемся алгоритмом 8

Алгоритм 8: Алгоритм поиска взвешенных совпадений для подграфа

Вход: подграф $G_i^j(V_i^j, E_i^j)$, списки соответствий ξ

Выход: новый список соответствий ξ_i^j

```

1 начало Matching
2    $\xi_i^j := []$ 
3   для всех конкатенаций всевозможных сочетаний кортежей в
      списках  $\xi(v')$ ,  $v' \in V_i^j$  выполнять
4     Пусть  $((q_0, v_0), \dots, (q_M, v_M))$  – соответствующая конкатенация.
5     если среди вершин  $v_0, \dots, v_M$  есть одинаковые тогда
6       Перейти к следующей итерации
7      $flag := True$ 
8     для всех дуг  $(q_s, q_r)$  графа  $G'$ , концами которых являются
      вершины  $q_0, \dots, q_M$  выполнять
9       если  $(v_s, v_r) \notin E \vee \rho(\chi(q_s, q_r), \chi'(v_s, v_r)) = 0$  тогда
10         $flag := False$ 
11     если  $flag = True$  тогда
12       Добавить соответствие  $((q_0, v_0), \dots, (q_M, v_M))$  в список  $\xi_i^j$ 
13 вернуть список  $\xi_i^j$ 

```

Таким образом, для каждого полученного на этапе раскраски подграфа G_i^j будет получен единственный список соответствий ξ_i^j . После этого нам остаётся описать переход к следующей итерации цикла.

Для следующей итерации мы строим граф $G_{i+1}(V_{i+1}, E_{i+1})$, для которого:

$$\begin{aligned}
 V_{i+1} &= \{G_i^1, \dots, G_i^K\}, \\
 E_{i+1} &= \{(G_i^{j'}, G_i^{j''}) | \exists v' \in V_i^{j'} \quad \exists v'' \in V_i^{j''} \quad \exists (v', v'') \in E_i\}, \\
 \xi(G_i^j) &= \xi_i^j, j = 1, 2, \dots, K
 \end{aligned}$$

В силу связности графа G' алгоритм завершается для любого лимита вершин $N > 1$. Поскольку выполняется сравнение характеристик всех рёбер (а по построению графа G_0 – и меток вершин), список соответствий $\xi(v^*)$, где v^* – единственная вершина графа, на котором произошёл выход из цикла, содержит списки точных вершинных соответствий, по каждому из которых можно восстановить взвешенное соответствие как подграф графа G . Действительно, если $((q_0, v_0), (q_1, v_1), \dots, (q_l, v_l))$ – некое вершинное совпадение, то граф $\widehat{G}(\widehat{V}, \widehat{E})$, где

$$\widehat{V} = \{v_0, v_1, \dots, v_l\},$$

$$\widehat{E} = \{(v_i, v_j) | (q_i, q_j) \in E'\},$$

есть взвешенное совпадение на архивном графе G .

5 Список литературы

1. Берж, К. Теория графов и ее применения. / Берж К.– М.: Издательство иностранной литературы, 1962.– 320 с.
2. Ерусалимский, Я.М. Дискретная математика. Теория, задачи, приложения: учеб. пособие / Я.М. Ерусалимский.– М.: Вузовская книга, 2009.– 288 с.
3. Лекции по теории графов. / Емеличев В. А. [и др.] – М.: Наука, 1990. – 384 с.
4. Мурсалиева, Г. Группы смерти (18+) [Электронный ресурс] / Г. Мурсалиева // Новая газета.– 2016.– Режим доступа: <https://www.novayagazeta.ru/articles/2016/05/16/68604-gruppy-smerti-18>.– (Дата обращения 05.06.2018).
5. Babai, L. Graph Isomorphism in Quasipolynomial Time. [Электронный ресурс] / L. Babai. // arXiv:1512.03547.– 2015.– 89 p.– Режим доступа: <https://arxiv.org/abs/1512.03547>.– (Дата обращения 08.06.2018).
6. Bellman, R. On a Routing Problem. / R. Bellman // Quarterly of Applied Mathematics.– 1958.– Vol 16, No. 1.– P. 87-90.
7. Imani, N., Sarbazi-Azad, H., G. Akl, S., Some topological properties of star graphs: The surface area and volume [Электронный ресурс] / N. Imani, H. Sarbazi-Azad, S. G. Akl // Discrete Mathematics. – 2008.– Режим доступа: <https://doi.org/10.1016/j.disc.2008.08.007>.– P. 560-569. – (Дата обращения 07.06.2018).
8. Loeliger, J., McCullough, M. Version Control with Git. / Jon Loeliger, Matthew McCullough.– Sebastopol: O'Reilly, 2012.– 436p.
9. Python 3.6.5 documentation [Электронный ресурс] // Режим доступа: <https://docs.python.org/3/index.html>.– (Дата обращения 05.06.2018).
10. Towards Practical and Robust Labeled Pattern Matching in Trillion-Edge Graphs. / T. Reza [и др.] // 2017 IEEE International Conference on Cluster Computing.– 2017.– 12 p.
11. Tutorial – NetworkX 2.1 documentation [Электронный ресурс] // Режим доступа: <https://networkx.github.io/documentation/stable/tutorial.html>. – (Дата обращения 04.06.2018).