

МЕТОД РЕШЕНИЯ ЗАДАЧИ РАЗЛИЧЕНИЯ ОРГРАФОВ НА ОСНОВЕ СЛОЖНОСТИ

В.А. Кохов,

кандидат технических наук, доцент кафедры Высшей математики
на факультете экономики Национального исследовательского университета
«Высшая школа экономики».

В.В. Кохов,

студент кафедры «Прикладная математика»
Московского энергетического института,
e-mail: viktorkokhov@rambler.ru.
Адрес: г. Москва, ул. Красноказарменная, д. 14.

Предложен метод решения задачи различения орграфов. Основу метода определяет матричная модель сложности, которая учитывает количественные и качественные характеристики фрагментов орграфа. Модель впервые позволяет оценивать значимость каждого фрагмента орграфа в его общей сложности. Приведены результаты решения задач различения и определения сходства для орграфов.

Ключевые слова: орграф, сложность орграфа, проблема различения орграфов, модель сложности, сходство орграфов, расположение фрагментов.

1. Введение

Сложность системы, будучи ключевым понятием системологии, требует глубоких и интенсивных исследований и ориентирована на развитие подходов к анализу сходства систем.

Сложность и сходство структурированных объектов (корпоративных социальных сетей; семантических сетей, абстрактных графов, графов синтаксической структуры предложений и др.) наиболее востребована в таких областях, как системы искусственного интеллекта, информационно-поисковые системы структурной информации (семантического *web*-поиска текстовых документов) и др. [1-3].

Приведем краткий обзор по построению моделей сложности графов. Одна из первых моделей была связана с построением теоретико-информационных индексов [4,5]. Пусть имеется граф $G=(V, E)$, где $|V|=p$, $|E|=q$ и множество его вершин V , разбивается на k классов эквивалентности V_i , $i=1..k$. В качестве отношения эквивалентности может выступать принадлежность вершин к одной орбите группы автоморфизмов графа. Тогда *информационное содержание* (количество информации), приходящееся на одну вершину, будет определяться следующей формулой:

$$ИИС(G/V) = - \sum_{i=1}^k C_i \log_2 C_i,$$

где $C_i = |V_i|/|V|$. Полное информационное содержание графа, рассчитанное на основе V , будет $ТИС(G/V) = |V|ТИС(G/V)$. $ТИС(G/V)$ рассматривается как интегральная количественная мера его структурной сложности (неоднородности). На основе информационного содержания графа может быть построено множество различных индексов сложности, если вместо множества вершин анализировать другие множества однотипных фрагментов, например цепей с длиной два (P_2). Из [6] известен орбитальный информационный индекс графа в базисе связей $\eta(P_2)$:

$$ОИ(G/P_2) = 2\eta \log_2 \eta - \sum_{i=1}^k \eta_i \log_2 \eta_i,$$

где k – число орбит группы расположения η в G , η_i – количество связей в i -ой орбите. По сути, это полное информационное содержание графа, рассчитанное на основе связей, плюс « $\eta \log_2 \eta$ » – информация о размере графа (число связей).

В [7] для определения сложности графа G введена функция

$$CM(G) = \frac{pq}{p+q} \sum_{i < j} \gamma(v_i, v_j),$$

где $\gamma(v_i, v_j)$ – число простых цепей из вершины v_i в вершину v_j . Указано, что эта функция соответствует интуитивному понятию сложности графа. В [8] предложена функция, значение которой, равно числу остовных деревьев (каркасов) графа. В качестве функции сложности в [9] используется число деревьев, изоморфно вкладываемых произвольным образом в граф G .

В [10] рассматривается топологическая сложность графовых моделей систем (ГМС), причем в наиболее общем и системном виде. Выделены девять требований, предъявляемых к моделям (функциям) сложности систем. В [11] приведен обзор подходов к определению структурной сложности графов и выделены три этапа в построении моделей: (1) построение индексов сложности, то есть числовых инвариантов графов, отражающих одну из сторон сложности, например сложность в терминах цепных, циклических, древесных и др. характеристик; (2) построение моделей сложности, основанных на мультииндексах, отражающих несколько сторон комплексного понятия «сложность»; (3) построение концептуальных и математических моделей, включающих многоуровневое отражение сложности, с привлечением достижений теории систем, теории графов, теории информации, топологии и

др. Показано, что к перспективным направлениям построения моделей сложности ГМС четвертого поколения следует отнести такие матричные модели, которые учитывают три уровня характеристики в единой модели: (1) сложность ГМС как единого целого; (2) сложность фрагментов ГМС; (3) сложность расположения фрагментов ГМС в ее топологии с характеристикой вкладов в общую сложность.

Ниже предлагается матричная модель сложности орграфов, позволяющая:

- ♦ учитывать как количественные, так и качественные характеристики систем, представленных орграфами;
- ♦ обобщить индексы симметрии расположения «связок» до индексов симметрии расположения любых фрагментов орграфа и вычислять их значения все с большей и большей точностью;
- ♦ проводить иерархический анализ сложности орграфов с уточнением результатов на каждом следующем уровне иерархии;
- ♦ определять вклад любого фрагмента в общую сложность;
- ♦ все более и более точно решать задачи различения орграфов;
- ♦ все более и более точно решать задачи различения расположения фрагментов в орграфах;
- ♦ формировать и исследовать новые виды отношений эквивалентности орграфов.

2. Основные определения

Пусть $G=(V,E)$ орграф. Два орграфа $G=(V,E)$ и $G^*=(V^*,E^*)$ изоморфны ($G \approx G^*$), если

$$\exists \varphi: V \leftrightarrow V^* \text{ и } \forall u, v \in V [(u,v) \in E \leftrightarrow (\varphi(u), \varphi(v)) \in E^*],$$

где $\varphi(u), \varphi(v) \in V^*$.

Орграф $G^*=(V^*,E^*)$ изоморфно вкладывается в орграф $G=(V,E)$ как порожденный подграф ($G^* \subseteq^s G$), (подграф ($G^* \subset G$)), если в G есть порожденный подграф (подграф) $G^*=(V^*,E^*)$, для которого справедливо отношение $G^* \approx G$.

Изоморфизм орграфа на себя называется автоморфизмом и обозначается через ξ . Множество всех автоморфизмов орграфа образует группу по умножению автоморфизмов и обозначается через $Aut(G)$. Порядок группы обозначим через $|Aut(G)|$. Под числом канонических изоморфных вложений G^* в G будем понимать величину, определяемую следующим образом $w(G^*, G) = W(G^*, G) / |Aut(G)|$, где $W(G^*, G)$ – число всех изоморфных вложений G^*

в G . Орбитой вершины $v \in V$ называется подмножество $\theta(Aut(G), v)$ вершин орграфа G , которые могут быть отображены на вершину v :

$$\theta(Aut(G), v) = \{v^* : \exists \xi \in Aut(G) : \xi(v^*) = v\}.$$

Орбиты вершин орграфа определяют классы их эквивалентного расположения.

Абстрактный тип t — произвольный орграф, определённый с точностью до изоморфизма. Группу его вершинных автоморфизмов обозначим через $Aut(t)$. Обозначим множество всех канонических изоморфных вложений абстрактного типа t в G через $F^t(G) = \{f_1^t, f_2^t, \dots, f_m^t\}$, а через $|F^t(G)|$ обозначим количество фрагментов типа t в G . Если на множестве вершин типа фрагмента t и G задана нумерация, то фрагмент f^t орграфа G может быть представлен помеченными фрагментами $f^{t'}$, когда каждой вершине типа фрагмента t сопоставляется номер вершины орграфа G , которой она соответствует при вложении. Число помеченных фрагментов, представляющих один и тот же фрагмент G , равно порядку группы вершинных автоморфизмов абстрактного типа.

Пусть $Aut^t(G)$ является индуцированным представлением группы $Aut(G)$ и определяет симметрию расположения фрагментов типа t в G . Под t -автоморфизмом G будем понимать подстановку ξ^t на множестве помеченных фрагментов типа t орграфа G (или канонических изоморфных вложений) $F^t(G)$, индуцированную некоторым вершинным автоморфизмом ξ орграфа G . В процессе индуцирования помеченный фрагмент $f_i^t \in F^t(G)$, представленный каноническим изоморфным вложением (v_1, v_2, \dots, v_n) орграфа G , переходит в помеченный фрагмент $f_j^t(G) \in F^t(G)$, каноническое представление которого получено канонизацией вложения (u_1, u_2, \dots, u_n) , где $u_i = \xi(v_i)$, $i = 1..n$, n — число вершин фрагмента типа t . Группой t -автоморфизмов G ($Aut^t(G)$ или t -группа) будет группа подстановок, носителем которой является всё множество t -автоморфизмов для данного t , а групповой операцией — операция произведения подстановок. Тот факт, что множество t -автоморфизмов образует группу, непосредственно следует из свойств $Aut(G)$. Степень t -группы равна числу канонических изоморфных вложений абстрактного типа t в орграф G ($|F^t(G)|$), а порядок меньше или равен порядку $Aut(G)$, так как два различных нетождественных вершинных автоморфизма могут индуцировать один и тот же t -автоморфизм. Все понятия, связанные с анализом $Aut^t(G)$, определяются аналогично понятиям, связан-

ными с анализом $Aut(G)$ (например, орбиты t -группы). t -группа точно характеризует симметрию расположения фрагментов типа t в орграфе G . Все неопределяемые ниже понятия можно найти в [12, 13].

3. Задачи различения орграфов и различения расположения их фрагментов

Пусть \mathfrak{X}_R обозначает множество всех связных орграфов, на котором задано отношение R — «быть изоморфными». Тогда задача различения (установления факта изоморфизма) двух орграфов $G_i, G_j \in \mathfrak{X}$ определяется следующими параметрами:

$$DG = \langle \mathfrak{X}; G_i, G_j; R \rangle.$$

Решением задачи DG является ответ на вопрос, являются ли G_i и G_j изоморфными?

Задача различения расположения двух фрагментов f_i^t, f_j^t типа t в орграфе G определяется следующими параметрами:

$$DF^t = \langle G; f_i^t, f_j^t; \xi^t \rangle,$$

где ξ^t — отношение принадлежности фрагмента к одной орбите t -группы. Решением задачи является ответ на вопрос, принадлежат ли f_i^t, f_j^t одной и той же орбите t -группы?

4. Инварианты и их применение для решения задач различения орграфов

Основным инструментом при решении задач различения орграфов является понятие инварианта орграфа. Пусть Q_τ — непустое множество с отношением эквивалентности τ (множество чисел, векторов, матриц, графов и т.д.). Функция IN , заданная на множестве $\mathfrak{X}_R(G)$ и принимающая значения в Q_τ , называется инвариантом орграфа G , если справедливо условие:

$$\forall G_i, G_j \in \mathfrak{X}[G_i(R)G_j \rightarrow IN(G_i)(\tau)IN(G_j)].$$

Графы G_i, G_j IN -эквивалентны $(G_i(IN)G_j)$, если $IN(G_i)(\tau)IN(G_j)$.

Инвариант IN называется полным инвариантом орграфа, если выполняется условие:

$$\forall G_i, G_j \in \mathfrak{X}[IN(G_i)(\tau)IN(G_j) \rightarrow G_i(R)G_j].$$

Под чувствительностью инварианта орграфа G в множестве $T \subseteq \mathcal{R}$ будем понимать отношение вида: $\alpha(IN, R) = |T(IN)| / |T(R)|$, где $|T(IN)|$ обозначает число классов IN -эквивалентности орграфов, а $|T(R)|$ – число неизоморфных орграфов в множестве T . Чувствительность инварианта определяет точность решения задачи различения орграфов в множестве T . Границей вырождаемости инварианта IN в множестве $T \subseteq \mathcal{R}$ называется наименьшее число вершин $G \in T$, при котором инвариант IN становится неполным.

5. Матрица достроек фрагментов как инвариант характеристики орграфа и характеристики расположения его фрагментов

Пусть $F^l(G) = \{F^1, F^2, \dots, F^l, \dots, F^T\}$ ($F^l(G) = \{F^{l1}, F^{l2}, F^{ln}, \dots, F^{lT}\}$) обозначает множество всех собственных фрагментов (помеченных фрагментов) орграфа G , а $F^u(G) = \{f_1^u, f_2^u, \dots, f_j^u, \dots, f_{rt}^u\}$ множество помеченных фрагментов типа t, j – номер фрагмента, rt – число фрагментов типа t, T – число типов фрагментов. Пусть $B = \langle b_1, b_2, \dots, b_j, \dots, b_{k1} \rangle$ обозначает базис структурных дескрипторов (СД), определяющий точность характеристики орграфа, и пусть $m_{i,j}$ – число достроек фрагмента f_j^u орграфа G до фрагмента орграфа, изоморфного $b_j \in B$. Построим матрицу $EM(G/B) = \|m_{i,j}\|$, где $i = 1..|F^l(G)|, j = 1..k1$. Выберем базис путей, как первоочередной для исследования, что обосновано следующими причинами: (1) любые орграфы включают пути; (2) пути отражают

как локальные (пути малой длины), так и глобальные (пути большой длины) свойства орграфов; (3) пути – наиболее легко конструктивно перечислимые фрагменты в сравнении с другими фрагментами орграфов; (4) пути относятся к классу монотонно нарастаемых по сложности базисов СД для характеристики орграфов.

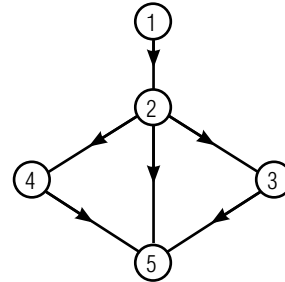


Рис. 1.

Для орграфа G (рис. 1) приведены примеры двух видов матриц $EM(G/V \subseteq P)$ (табл. 1, выделены серым цветом), в которых в качестве строк выступают вершины G , а в качестве базиса СД – базис путей $P = \langle P_0, P_1, \dots, P_{(p-1)} \rangle$. Матрица, упорядоченная лексикографически по убыванию значений строк (например, справа налево) является инвариантом, если не сохраняется нумерация вершин G . В первом виде матриц достройка выполняется от вершин как конечных в пути, а во втором – достраиваемый путь включает вершину в любом месте пути. Учитывая, что пути можно достраивать до путей-подграфов (P) или путей порожденных подграфов (P^S), получаем четыре вида основных матриц достроек путей до путей:

Таблица 1.

Вершина конечная достраивается до путей						
V	P_0	P_1	P_2	P_3	irc ($f^l(C.n)$)	irc ($f^l(C)$)
5	1	3	3	2	0,3846	0,3846
1	1	1	3	2	0,3616	0,3616
2	1	4	2	0	0,1231	0,1308
3	1	2	1	0	0,0654	
4	1	2	1	0	0,0654	
Slw	5	12	10	4	31	
Sw	1	2	2	2	7	
WF	5	6	5	2	18	
ISC	5	18	45	62	$ISC(G/P)=130$	

Вершина достраивается до путей						
V	P_0	P_1	P_2	P_3	irc ($f^l(C.n)$)	irc ($f^l(C)$)
2	1	4	5	2	0,2885	0,2885
5	1	3	3	2	0,2308	0,2308
1	1	1	3	2	0,2077	0,2077
3	1	2	2	1	0,1365	0,2731
4	1	2	2	1	0,1365	
Slw	5	12	15	8	40	1
Sw	1	2	3	4	10	
WF	5	6	5	2	18	
ISC	5	18	45	62	$ISC(G/P)=130$	

$$EM(G / P^l \subseteq^e P); EM(G / P^l \subseteq P);$$

$$EM(G / P^l \subseteq^{Se} P); EM(G / P^l \subseteq^S P).$$

В наиболее общем виде матрица $EM(G / F^l \subseteq F)$ в качестве строк включает все подграфы из множества $F^l(G)$, а в качестве столбцов – все подграфы из множества $F(G)$. Таким образом, получаем стратифицированную систему матриц от $EM(G / V \subseteq P)$ до $EM(G / F^l \subseteq F)$, расширяемую по трем направлениям: (1) от V до F^l ; (2) от P_0 до F ; (3) одновременно по (1) и (2). Это приводит к формированию и исследованию *новых систем уточняемых отношений эквивалентности орграфов*.

Чтобы упорядочить фрагменты из $F(G)$ по возрастанию сложности, определим индекс сложности орграфа. Под структурным спектром G в базе СД $B = \langle b_1, b_2, \dots, b_p, \dots, b_{k1} \rangle$ будем понимать запись следующего вида $SS(G/B) = \beta_1 b_1, \beta_2 b_2, \dots, \beta_i b_i, \beta_{k1} b_{k1}$, где $\beta_i = 1$, если фрагмент $b_i \in B$ изоморфно вкладывается в G и $\beta_i = 0$ в противном случае [14].

Пусть для G построен его *полный структурный спектр* (ПСС) в базе СД

$$B = \langle b_1, b_2, \dots, b_p, \dots, b_{k1} \rangle; FS(G/B) =$$

$$= \langle w_1 b_1, w_2 b_2, \dots, w_i b_i, \dots, w_{k1} b_{k1} \rangle,$$

где b_i – фрагмент базиса; w_i – число канонических изоморфных вложений фрагмента b_i в G . Очевидно, что $w(K_1) = p$, а $w(K_2) = q$. Примем $ISC(K_1) = 1$, $ISC(K_2) = 3$. Так как для любого фрагмента f_i орграфа G можно определить его ПСС, а для каждого фрагмента от фрагмента G , можно построить его ПСС и т.д., то рекурсивным методом всегда можно вычислить *индекс спектральной сложности (ISC)* (ниже сложности) и вектор-индекс сложности (V_ISC) орграфа G в базе СД B :

$$ISC(G/B) = w_1 ISC(b_1) +$$

$$+ w_2 ISC(b_2), \dots, w_j ISC(b_j), \dots, w_{k1} ISC(b_{k1});$$

$$V_ISC(G/B) = \langle w_1 ISC(b_1); w_2 ISC(b_2); \dots;$$

$$w_j ISC(b_j); \dots; w_{k1} ISC(b_{k1}) \rangle.$$

Для G (рис. 1) при использовании базиса путей-подграфов $P = \langle P_0, P_1, P_2, P_3 \rangle$ получим:

- ◆ $ISC(G/P) = 5 \cdot 1 + 6 \cdot 3 + 5 \cdot 9 + 2 \cdot 31 = 130$;
- ◆ $V_ISC(G/P) = \langle 5; 18; 45; 62 \rangle$;
- ◆ $SS(G/P) = \langle 1, 1, 1, 1 \rangle$;
- ◆ $FS(G/P) = \langle 5, 6, 5, 2 \rangle$.

6. Расширенная матрица достроек фрагментов и матрица вкладов фрагментов в общую сложность орграфа

Эта матрица определяет основу иерархического метода анализа сложности орграфа с вычислением вкладов фрагментов, вкладов каждого класса эквивалентно расположенных фрагментов для каждого типа t и вкладов фрагментов в сложность орграфа.

Добавим к матрице $EM(G / F^l \subseteq B)$ пять новых строк:

1. Строки с номером нуль для вектор-индекса сложности элементов $b_j \in B$:

$$V_ISC(G/B) = \langle ISC(b_1), \dots, ISC(b_j), \dots, ISC(b_{k1}) \rangle.$$

2. Строки с номером $(k+1)$ со значениями элементов

$$Slw(F^l / b_j) = \sum_{f^t \in F^l} \sum_{i=1}^{rt} m_{ij}.$$

3. Строки с номером $(k+2)$, значения элементов которой $Sw(b_j)$, равны суммарному числу канонических изоморфных вложений фрагмента типа t в $b_j \in B$ по всем типам $t = 1, 2, \dots, T$, т.е.

$$Sw(F^l / B) = \langle Sw(F^l / b_1), \dots, Sw(F^l / b_j), \dots, Sw(F^l / b_{k1}) \rangle,$$

$$\text{где } Sw(F^l / b_j) = \sum_{t=1}^T \frac{f^t(b_j)}{|Aut(f^t)|} = \sum_{t=1}^T w(f^t / b_j).$$

4. Строки с номером $(k+3)$ со значениями элементов

$$w_j(b_j) = Slw(F^l / b_j) / Sw(F^l / b_j).$$

5. Строки с номером $(k+4)$ со значениями элементов

$$V_ISC(G / b_j) = w_j(b_j) ISC(b_j).$$

В результате построим расширенную матрицу достроек фрагментов

$$EM^*(G / F^l \subseteq B).$$

На основе $EM^*(G / F^l \subseteq B)$ построим матрицу

$$MIRC(F^l \subseteq B(G)) = \left\| irc(f_i^t / b_j) \right\|, \quad i = 0..k+4; j = 1..k1+3$$

относительных вкладов фрагментов в сложность G.

Значения элементов матрицы вычисляются по формуле

$$irc(f_i^t / b_j) = \frac{m_{ij}}{Sw(F^t / b_j)} \frac{ISC(b_j)}{ISC(G / B)}.$$

Тогда величина $irc(f_i^t / B)$, вычисляемая по формуле

$$irc(f_i^t / B) = \frac{1}{ISC(G / B)} \sum_{j=1}^{k1} m_{ij} \frac{ISC(b_j)}{Sw(F^t / b_j)},$$

определяет относительный вклад f_i^t в общую сложность при использовании базиса B .

Фрагменты f_i^t , имеющие одинаковые значения вкладов $irc(f_i^t / B)$, образуют класс $f^t(C)$ эквивалентных по расположению фрагментов типа t с общим вкладом $irc(f^t(C) / B)$. Сумма относительных вкладов по всем фрагментам одного типа t образует вклад $irc(f^t / B)$. Расширенную матрицу достроек фрагментов, дополненную тремя столбцами:

- ◆ $(k1+1)$ со значениями $irc(f^t(C.n) / B)$ для каждого типа t ;
 - ◆ $(k1+2)$ со значениями $irc(f^t(C) / B)$ для каждого типа t ;
 - ◆ $(k1+3)$ со значениями $irc(f^t / B)$ для каждого типа t ,
- обозначим $MIRC(F^t \subseteq B(G))$.

Пример матрицы $MIRC(V \subseteq P(G))$ без значений $irc(f^t(C.n) / b_j)$ для орграфа G (рис. 1) приведен в табл. 1. На основе $MIRC(F^t \subseteq F(G))$ легко построить матрицу абсолютных вкладов фрагментов в сложность G , т.е. матрицу $MIAC(F^t \subseteq F(G))$.

7. Метод иерархического анализа сложности орграфов

Вычисление матриц $MIRC(F^t \subseteq B(G))$ позволяет проводить иерархический анализ сложности, включающий пять основных уровней, с последовательным уточнением результатов на основе:

1. Индексов сложности – $ISC(G/B)$;
2. Вектор-индексов сложности – $V_ISC(G/B)$;
3. Вектор-индексов относительных вкладов каждого типа фрагментов t (вершины, дуги, пути с длиной 2 и т.д.) – $V_irc(G, F^t)$, где $t = 1..T$;
4. Вектор-индексов относительных вкладов классов фрагментов в каждом типе фрагментов

$$V_irc(G, F^t(C)),$$

где C – номер класса фрагментов типа t и $t = 1..T$;

5. Вектор-индексов относительных вкладов каждого фрагмента с номером n в классе C типа t –

$$V_irc(F^t(C.n)/B), \text{ где } t = 1..T.$$

Дополнительные уровни анализа сложности определяются на основе вектор-индексов относительных вкладов каждого фрагмента с номером n в классе C типа t по каждому элементу $b_j \in B$ –

$$V_irc(F^t(C.n)/b_j), \text{ где } t=1..T, j=k1, (k1-1), \dots, 1.$$

Для трех орграфов (рис. 2) в табл. 2 приведены матрицы $MIAC(P^t \subseteq P(G))$, упорядоченные по убыванию значений вкладов всех путей по каждому типу t , где $t=P_0 - P_2$.

Приведем результаты сравнительного анализа трех орграфов по иерархической схеме:

- ◆ на первом уровне

$$ISC(G_1/P) = ISC(G_2/P) = ISC(G_3/P);$$

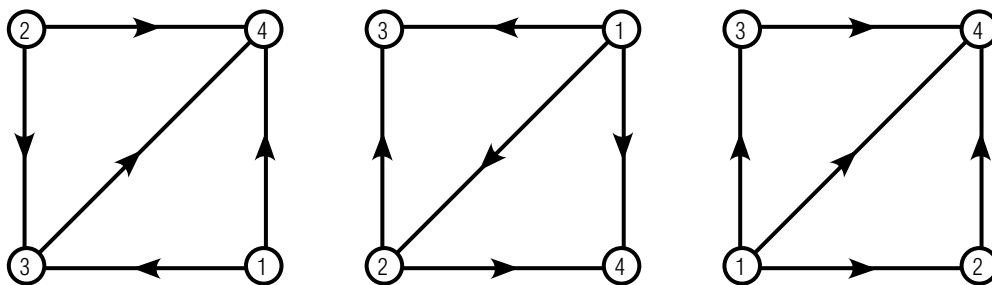


Рис. 2.

Таблица 2.

$MIAC(P^l \subseteq P(G_1)) = MIAC(P^l \subseteq P(G_2))$							$MIAC(P^l \subseteq P(G_3))$								
$P^l(G_1)$	$P^l(G_2)$	1	3	9	$irc(P^l(C.n))$	$irc(P^l(C))$	$irc(P^l)$	$P^l(G_3)$	1	3	9	$irc(P^l(C.n))$	$irc(P^l(C))$	$irc(P^l)$	
3	2	1	3	3	7	14	23	4	1	3	3	7	14	23	
4	1	1	3	3	7			1	1	3	3	7			
1	4	1	2	1,5	4,5	9		3	1	2	1,5	4,5	9		
2	3	1	2	1,5	4,5			2	1	2	1,5	4,5			
34	12	0	1	3	4	4	11	13	0	1	1,5	2,5	10	11	
13	23	0	1	1,5	2,5	5		34	0	1	1,5	2,5			
23	24	0	1	1,5	2,5			2	12	0	1	1,5			2,5
14	13	0	1	0	1	24			0	1	1,5	2,5			
24	14	0	1	0	1	2	14	0	1	0	1	1	1		
134	124	0	0	1,5	1,5	3	3	124	0	0	1,5	1,5	3	3	
234	123	0	0	1,5	1,5			134	0	0	1,5	1,5			
$Slw(b_j)$		4	15	18	37	37	37	$Slw(b_j)$		4	15	18	37	37	37
$Sw(b_j)$		1	3	9	10	$ISC(G/P)=37$		$Sw(b_j)$		1	3	9	10	$ISC(G/P)=37$	
$WF(b_j)$		4	5	2	11			$WF(b_j)$		4	5	2	11		
$V_ISC(b_j)$		4	15	18	37			$V_ISC(b_j)$		4	15	18	37		

◆ на втором уровне

$$V_ISC(G_1/P) = V_ISC(G_2/P) = V_ISC(G_3/P);$$

◆ на третьем уровне

$$V_irc(G_1, P^l(C)) = V_irc(G_2, P^l(C)) = V_irc(G_3, P^l(C));$$

◆ на четвертом уровне

$$V_irc(G_1, P^l(C.n)/B) = V_irc(G_1, P^l(C.n)/B) \neq V_irc(G_1, P^l(C.n)/B).$$

Таким образом, на четвертом уровне выявлено различие по сложности орграфов G_1 и G_2 с G_3 . Для G_1 и G_2 значения матриц $MIAC(P^l \subseteq P(G))$ одинаковые, следовательно, G_1 и G_2 IN-эквивалентные орграфы по $MIAC(P^l \subseteq P(G))$. Заметим, что полученные классы эквивалентно расположенных путей в $G_1 - G_3$, являются орбитами групп

$$Aut^{P_0}(G), Aut^{P_1}(G), Aut^{P_2}(G).$$

Если базис путей является достаточным для того, чтобы классы эквивалентно расположенных путей стали орбитами групп

$$Aut^{P_0}(G), Aut^{P_1}(G), \dots, Aut^{P_{(p-1)}}(G),$$

то нетрудно доказать, что справедливы следующие утверждения.

Утверждение 1. Индекс $ТИС(G/V)$ и индексы $ТИС(G/P_1), \dots, ТИС(G/P_{(p-1)})$ являются восстанавливаемыми характеристиками из $MIRC(P^l \subseteq P(G))$ для орграфа G .

Действительно, например, для орграфа G_1 (рис. 2) имеем

$$ТИС(G_1 / V) = WF(b_1) \left(\frac{|P^0(C_1)|}{WF(b_1)} \log_2(|P^0(C_1)|) + \frac{|P^0(C_2)|}{WF(b_1)} \log_2(|P^0(C_2)|) \right).$$

Утверждение 2. Индекс $ОИС(G/P_2)$ и индексы $ОИС(G/P_3), \dots, ОИС(G/P_{(p-1)})$ являются восстанавливаемыми характеристиками из $MIAC(P^l \subseteq P(G))$ для орграфа G .

Действительно, например, для орграфа G_1 (рис. 2) имеем

$$ОП(G / P_2) = 2WF(b_2) (|P^2(C_1)| \log_2(|P^2(C_1)|) + |P^2(C_2)| \log_2(|P^2(C_2)|) + |P^2(C_3)| \log_2(|P^2(C_3)|)).$$

Утверждение 3. Индекс $CM(G)$ является восстанавливаемой характеристикой из $MIRC(P' \subseteq P(G))$ для орграфа G .

Действительно, например, для орграфа G_1 (рис. 2) имеем

$$ОП(G_1 / P_2) = \frac{WF(b_1)WF(b_2)}{WF(b_1) + WF(b_2)} (WF(b_2) + WF(b_3)).$$

Если классы расположения путей не являются орбитами, то мы имеем метод *приближенного вычисления* рассмотренных индексов. Для точного вычисления индексов необходимо базис путей наращивать элементами типа полупути, контура, ордеревья и т.д. Выделим, что добавляя к базису путей, например, один полупуть вида $K_{1,2}$, мы построим $MIRC(P' \subseteq (P \cup K_{1,2})(G))$ и определим не изоморфизм G_1 и G_2 (рис. 2).

8. Результаты определения точности решения задач различения орграфов

На рис. 3 приведены диаграммы всех орграфов с $p=3$.

Сравнивая орграфы по значениям относительных вкладов путей в сложность ($irc(P' \subseteq C.n)$) из матриц

$MIRC(P' \subseteq P(G))$, получим: 10 классов:

$$(G_1) > (G_2) > (G_3) > (G_4) > (G_5) > (G_6, G_7) > (G_8, G_9) > (G_{10}) > (G_{11}) > (G_{12}, G_{13}).$$

Следовательно, точность решения задачи различения орграфов

$$\alpha(MIRC(P' \subseteq P(G)), R) = 0,769.$$

Заметим, что все пары орграфов, образующие классы эквивалентности, имеют инверсную ориентацию дуг. Такие орграфы будем называть *инверсными*. Для них справедливо

Утверждение 4. Неизоморфные инверсные пары орграфов являются *IN*-эквивалентными по значениям матрицы $MIRC(P' \subseteq P(G))$.

Следовательно, справедливо

Утверждение 5. Матрица $MIRC(P' \subseteq P(G))$ не является полным инвариантом орграфов и граница ее вырождаемости равна трем.

В табл. 3 приведены результаты определения точности решения задач различения для орграфов, сетей, графов по $ISC(G/P)$ и $MIRC(P' \subseteq P(G))$.

Использованы следующие обозначения: ND, NS, NG – число соответственно орграфов, сетей, графов; $\alpha(D, ISC), \alpha(S, ISC), \alpha(G, ISC)$ – чувствительность по ISC соответственно для орграфов, сетей, графов; $\alpha(D, MIRC), \alpha(S, MIRC), \alpha(G, MIRC)$ – чувствительность по $MIRC$ соответственно для орграфов, сетей, графов.

Анализ результатов приводит к следующим выводам: (1) граница вырождаемости $ISC(G/P)$ для

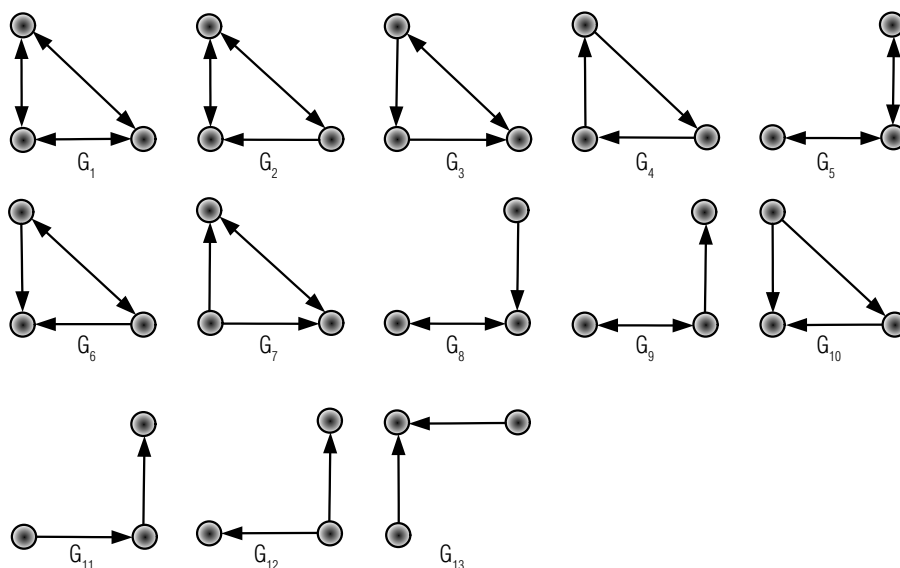


Рис. 3.

Таблица 3.

p	ND	$\alpha(D, ISC)$	$\alpha(D, MIRC)$	NS	$\alpha(S, ISC)$	$\alpha(S, MIRC)$	NG	$\alpha(G, ISC)$	$\alpha(G, MIRC)$
3	13	0,615	0,769	4	0,750	0,75	2	1	1
4	199	0,407	0,638	24	0,542	0,64	6	1	1
5	9364	0,183	0,533	267	0,269	0,51	21	0,947	1
6	1530843	0,064	0,596	5647	0,094	0,47	112	0,938	1
7				237317	0,019	0,47	852	0,977	1
8							11117	0,983	1
9							261080	0,989	1

орграфов (сетей, графов) равна 3 (3, 5); (2) граница вырождаемости $MIRC(P^l \subseteq P(G))$ для орграфов (сетей, графов) равна 3 (3, больше 9); (3) каждый класс эквивалентности орграфов включает инверсные пары орграфов; (4) для более точного решения задач различения орграфов (сетей) необходимо базис путей расширять полупутями.

9. Новые методы анализа сходства орграфов

Задача вычисления сходства орграфов включает следующие параметры:

- $M = \{G_1, G_2, \dots, G_p, \dots, G_n\}$ – множество орграфов, анализируемых на сходство.
- $B = \langle b_1, b_2, \dots, b_p, \dots, b_k \rangle$ – базис СД или правила его определения.
- $ISC(B) = \langle ISC(b_1); ISC(b_2); \dots; ISC(b_j); ISC(b_{k1}) \rangle$ – набор весов СД для учета их качественных характеристик.
- $D(G_i, G_j)$ – метрика (псевдометрика) или функция для вычисления коэффициентов сходства (несходства) (SC).

Необходимо построить матрицу (граф) попарных расстояний или коэффициентов сходства (несходства).

Выделим два метода решения задачи на основе матричных моделей:

- Метод иерархического уточняющего анализа, использующий в качестве метрик:

- ♦ на первом уровне – $|ISC(G_i / B) - ISC(G_j / B)|$;
- ♦ на 2-5 уровнях – метрику Евклида для соответствующих вектор-индексов вкладов из $MIRC(F^l \subseteq B(G))$, для каждого класса, полученного на предыдущем уровне.

- Метод анализа на основе подструктурного подхода (ПП), использующего вычисление попарных расстояний

$$D(G_i, G_j) = |V(F^l \subseteq B(G_i))| + |E(F^l \subseteq B(G_i))| + |V(F^l \subseteq B(G_j))| + |E(F^l \subseteq B(G_j))| - 2|V(mcf(F^l \subseteq B(G), F^l \subseteq B(G_j)))|,$$

где $F^l \subseteq B(G)$ – двудольная граф-модель с весами на дугах, соответствующая $EM(G/F^l \subseteq F)$, $mcf(F^l \subseteq B(G), F^l \subseteq B(G_j))$ – максимальный общий подграф по числу дуг для двух граф-моделей $F^l \subseteq B(G)$, $F^l \subseteq B(G_j)$ или коэффициент сходства

$$MSI(G_i, G_j) = (|V(mcf(F^l \subseteq B(G_i), F^l \subseteq B(G_j)))| + |E(mcf(F^l \subseteq B(G_i), F^l \subseteq B(G_j)))|^2) / (|V(F^l \subseteq B(G_i))| + |E(F^l \subseteq B(G_i))|) \cdot (|V(F^l \subseteq B(G_j))| + |E(F^l \subseteq B(G_j))|).$$

Приведем результаты анализа сходства на основе $EM(G/P^l \subseteq S F)$ по второму подходу для 13 орграфов (рис. 3) и результаты их кластеризации. В графе попарных расстояний получено 19 разных значений расстояний (весов ребер) в диапазоне от 0 до 48. При кластеризации из графа попарных расстояний последовательно удалялись ребра, значения веса которых, превышали соответственно следующие границы 9; 8; 6; 3; 2; 0. В результате выделено 7 одноэлементных кластеров и 3 двухэлементных, состоящих из неизоморфных инверсных орграфов (рис. 4). Такие же кластеры выделены на этапе 4 по методу подхода 1.

Заметим, что использование ПП, основанного на определении mcf для всех пар орграфов при одинаковом числе вершин, не позволяет проводить кластеризацию, из-за недостаточного диапазона значений весов ребер в графах попарного сходства.

Выделим, что, используя метрику Евклида для вычисления попарных расстояний между строками ма-

трицы $MIRC(F^I \subseteq F(G))$, мы решаем задачу определения попарного сходства *расположения фрагментов орграфа*.

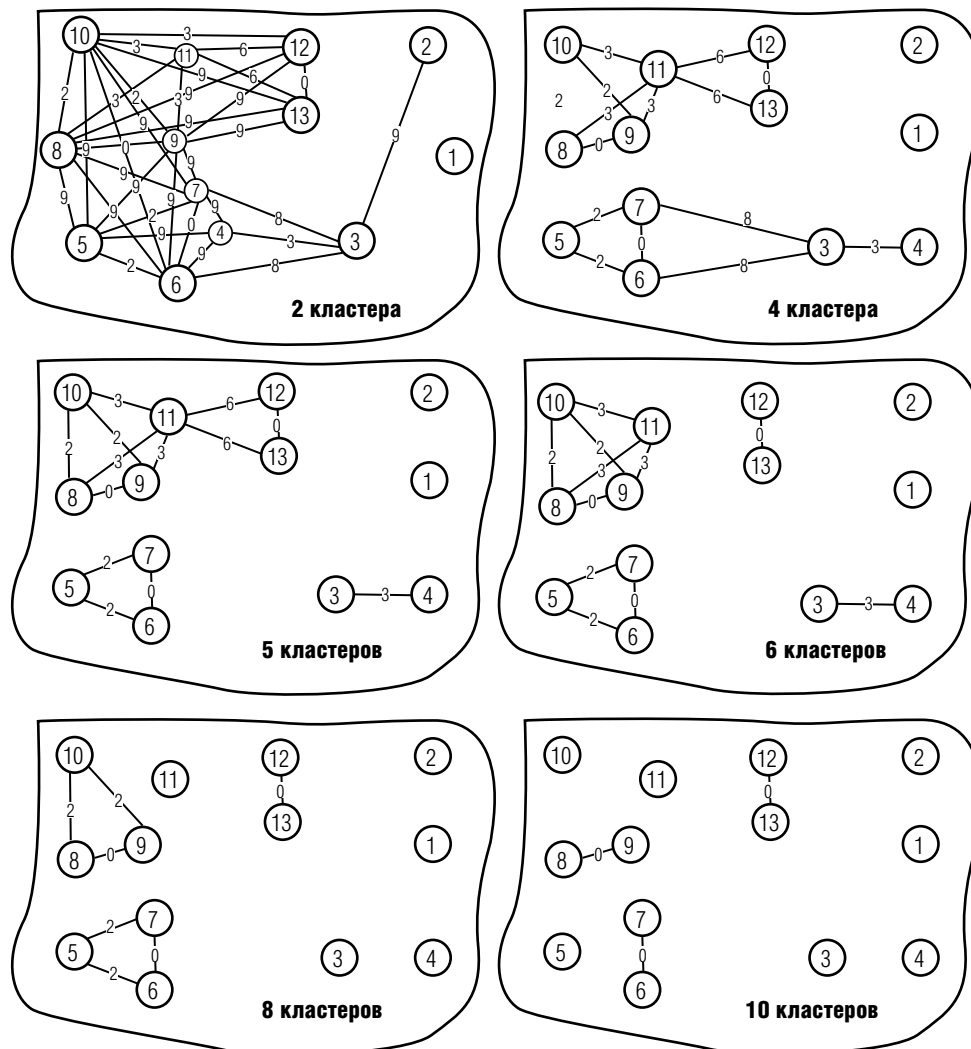


Рис. 4.

10. Экспериментальные оценки вычислительной сложности задачи построения матрицы относительных вкладов фрагментов в сложность орграфов

Экспериментальные оценки вычислительной сложности (ЭОБС) решения задачи построения матрицы $MIRC(P^I \subseteq P(G))$ получены на основе использования новых программных подсистем, включенных в АСНИ «Мастерская граф-моделей» [16, 17]. Эксперименты проводились на ноутбуке (SONY, Intel Core i5, 2.67 GHz, память 4 Gb, Windows 7).

Учитывая особенность ордеревьев, связанную с существованием для каждой пары вершин только одного пути, нетрудно получить теоретическую

полиномиальную оценку $T=O(p^3)$ вычислительной сложности алгоритма построения $MIRC(P^I \subseteq P(G))$ для ордеревьев. Основу алгоритма составляет метод достройки вершины ордерова до пути, заданной длины. На рис. 5 приведены графики ЭОБС алгоритма построения $MIRC(P^I \subseteq P(G))$ для ордеревьев со средними значениями индексов сложности ($ISC(G/P)$) и числами вершин от 108 до 508.

В общем случае задача построения $MIRC(P^I \subseteq P(G))$ для орграфов (сетей) является NP-полной проблемой, так как число путей всех длин в орграфах (сетях) растет экспоненциально и частным случаем для нее является классическая NP-полная проблема определения гамильтонова пути в орграфе (сети)

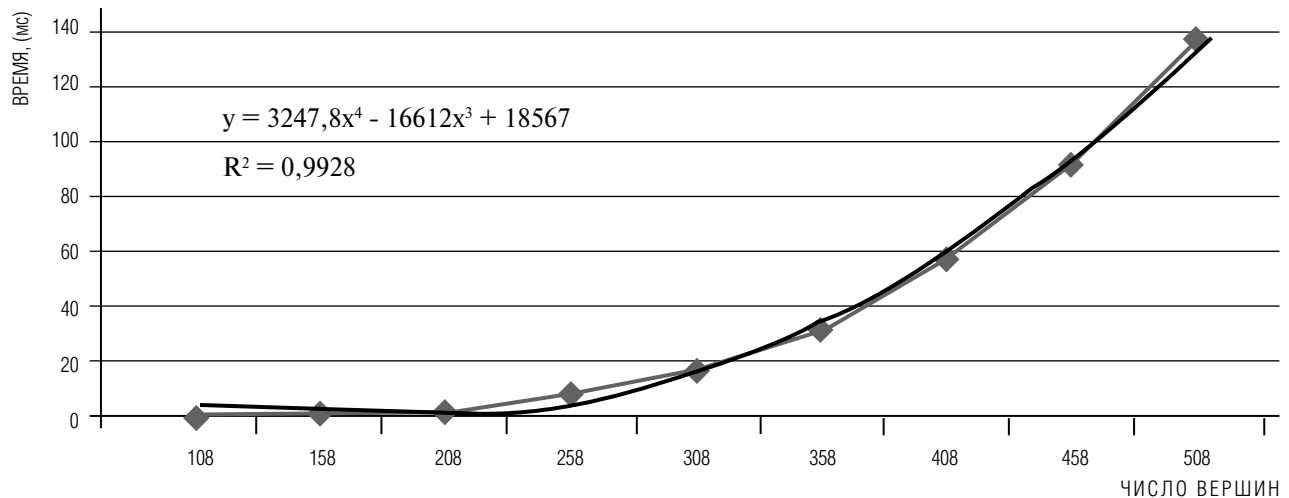


Рис. 5.

[15]. Сокращение перебора при построении матрицы $MIRC(P^i \subseteq P(G))$ методом достройки вершин до путей, заданной длины, осуществляется за счет учета симметрии (орбит стабилизаторов $Aut(G)$) и инвариантов вершин орграфов [13].

На рис. 6-9 приведены ЭОВС алгоритма построения восьми видов матриц:

$$MIRC(P^i_{0-0} \subseteq {}^S P_{0-12}(G)); MIRC(P^i_{0-2} \subseteq {}^S P_{0-12}(G);$$

$$MIRC(P^i_{0-4} \subseteq {}^S P_{0-12}(G)); MIRC(P^i_{0-6} \subseteq {}^S P_{0-12}(G);$$

$$MIRC(P^i_{0-0} \subseteq {}^S P_{0-20}(G)); MIRC(P^i_{0-2} \subseteq {}^S P_{0-20}(G);$$

$$MIRC(P^i_{0-4} \subseteq {}^S P_{0-20}(G)); MIRC(P^i_{0-6} \subseteq {}^S P_{0-20}(G);$$

Анализировались случайные орграфы с числом вершин от 100 до 600 и средней степенью (суммой полустепеней исхода и захода) равной четы-

рем. Были построены матрицы $MIRC$ следующих размеров (число строк \times число столбцов): 1) y_1 от 100×13 до 520×13 ; 2) y_2 от 622×13 до 3395×13 ; 3) y_3 от 1962×13 до 12369×13 ; 4) y_4 от 4945×13 до 39675×13 ; 5) x_1 от 100×21 до 520×21 ; 6) x_2 от 622×21 до 3395×21 ; 7) x_3 от 1962×21 до 12369×21 ; 8) x_4 от 4945×21 до 39675×21 .

Необходимо выделить, что задача вычисления сходства пары орграфов на основе поиска МОФ их матриц $MIRC(P^i \subseteq P(G))$ имеет полиномиальную вычислительную сложность при условии, что используется один и тот же базис P . Это обосновано возможностью сведения задачи поиска МОФ двух матриц $MIRC(P^i \subseteq P(G))$ к задаче поиска максимального паросочетания в двудольном орграфе с весами на дугах.

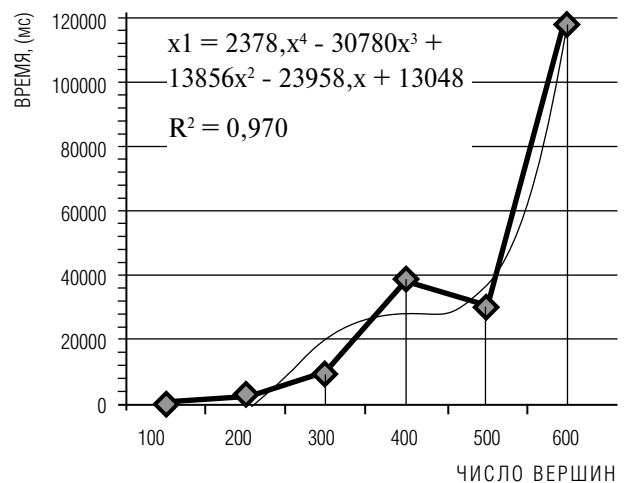
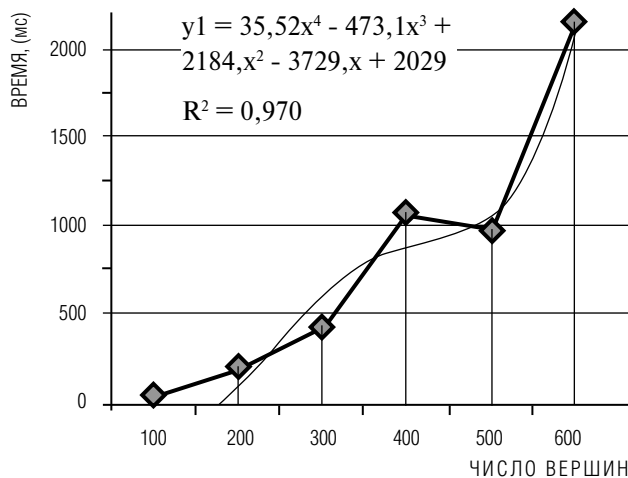


Рис. 6.

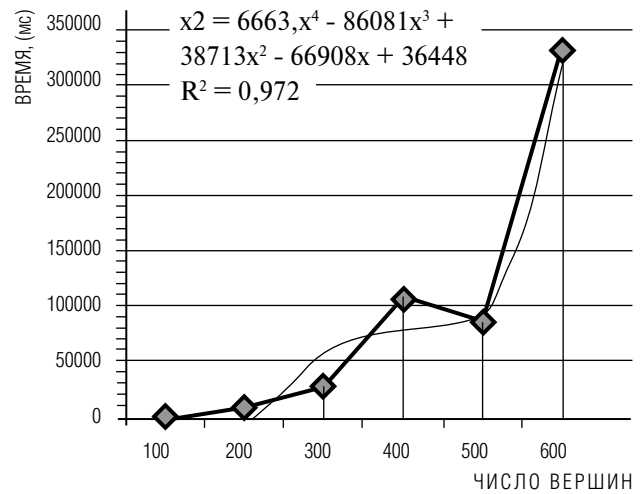
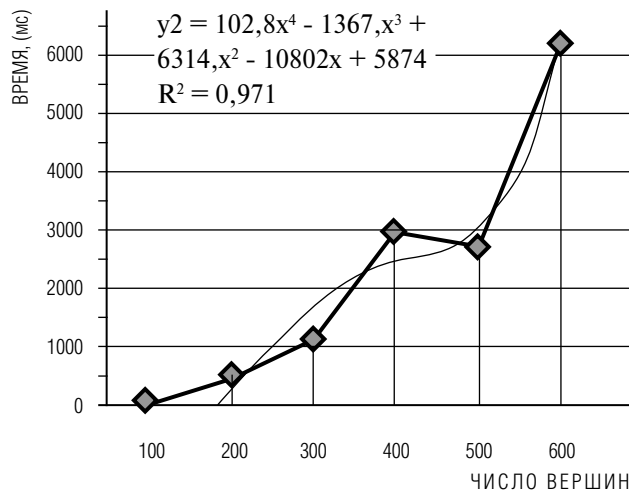


Рис. 7.

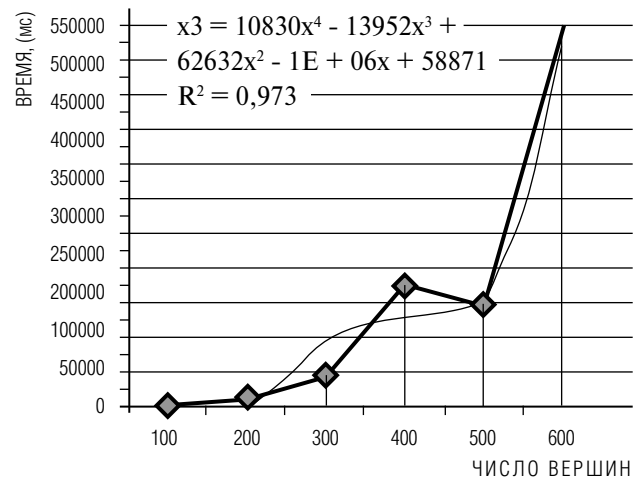
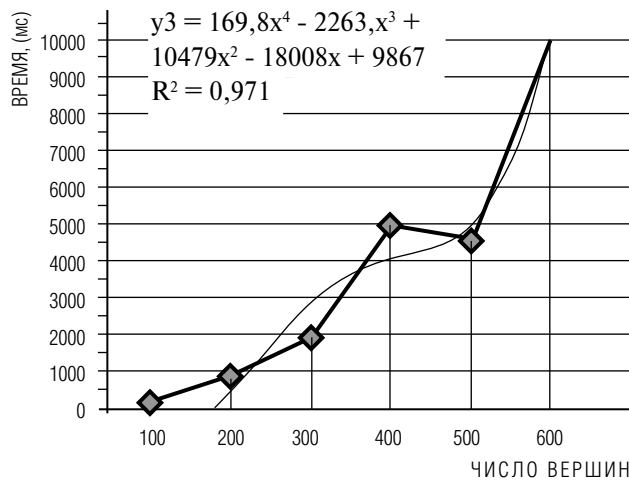


Рис. 8.

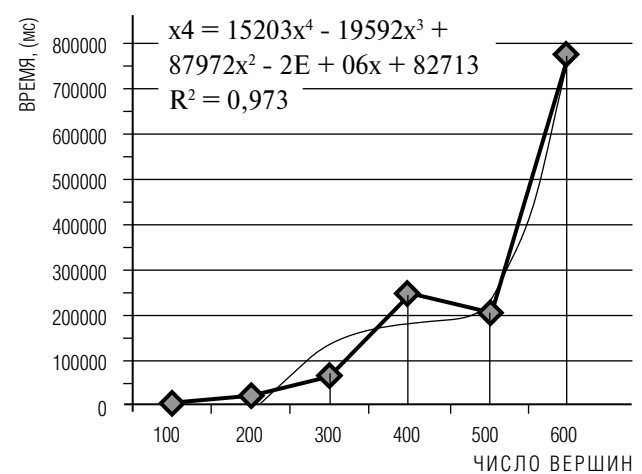
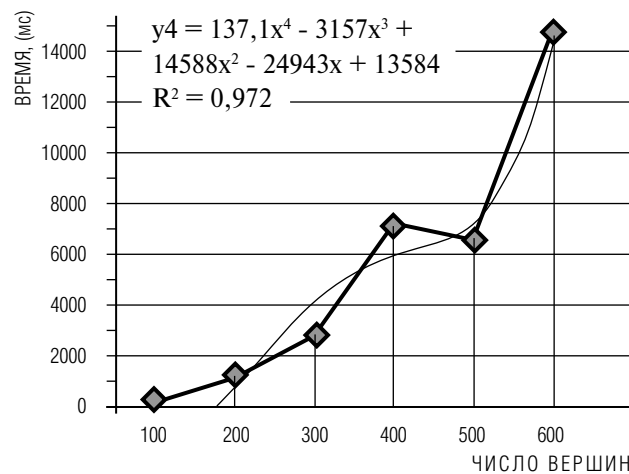


Рис. 9.

Заключение

Предложены матричные модели, позволяющие проводить иерархический уточняющий анализ сложности и сходства с учетом количественных и качественных характеристик фрагментов орграфов. Эти модели позволили расширить и дополнить теоретико-информационный подход к определению сложности орграфов качественными характеристиками его фрагментов. Рассмотрено применение моделей для решения теоретических

задач, связанных с определением границ вырождаемости матричных моделей в базисе путей для орграфов, сетей и графов. Модели полезны для решения прикладных задач различения и определения сходства ГМС, например, логико-вычислительных сетей, представляющих знания, сетей коммуникаций, компьютерных сетей и др. Предложенные методы анализа сложности и сходства программно реализованы и используются в учебном процессе НИУ ВШЭ, МЭИ (ТУ). ■

Литература

1. Финн В.К. Индуктивный метод соединенного сходства-различия и процедурная семантика ДСМ-метода. // НТИ, сер. 2, №4, 2010.
2. Кузнецов С.О. Финн В.К. Распространение процедур ЭС типа ДСМ на графы. // Техн. кибернетика, 1988, №5, С.4-11.
3. Игнатов Д.И., Кузнецов С.О. О поиске сходства интернет-документов с помощью частых замкнутых множеств признаков. Десятая национальная конференция по искусственному интеллекту с международным участием. КИИ-2006: Труды конференции. В 3-х т. М.: Физматлит, 2006. С. 245-249.
4. Шеннон К. Работы по теории информации и кибернетике. — М.: Изд. ин. лит., 1963.
5. Бончев Д.Г. Характеризация химических структур с помощью теории информации и теории графов. Автореферат дисс. докт. хим. наук. — Бургас, 1983. — 48 с.
6. Бертц С. Математическая модель молекулярной сложности. Химические приложения топологии и теории графов / Под ред. Р. Кинга, М.: Мир, 1987. — С. 236-258.
7. Minoly D. Combinatorial graph complexity. // Atti. Acad. Waz. Li. Rend. A. Sci. fis. mat. l'natur, vol. 59, no. 6. pp. 154-171, 1975.
8. Grone R., Herris R. A Bound for the Complexity of a Simple Graph. // Discrete Math., vol. 69, no. 1, pp. 97-99, 1988.
9. Загорянская А.А., Кохов В.А. Структурная сложность информационных систем и ее характеристика в базисе деревьев. Труды НТК студентов и аспирантов вузов России, Т1, Москва, 1998. — С.180-182.
10. Bonchev D., Polansky O.E. On the Topological Complexity on Chemical Systems // Studies in Physical and Theoretical Chemistry, 1987.V.51. P. 126-158.
11. Кохов В.А. Концептуальные и математические модели сложности графов. М.: Издательство МЭИ, 2002. — 160 с.
12. Касьянов В.Н., Евстигнеев В.А. Графы в программировании: обработка, визуализация и применение. — СПб.: БХВ-Петербург, 2003. — 1104 с.
13. Алгоритмы и программы решения задач на графах и сетях // Нечепуренко М.И., Попков В.К., Кохов В.А. и др. Новосибирск: Наука. 1990. — 515 с.
14. Кохов В.А. Метод количественного определения сходства графов на основе структурных спектров. Известия АН СССР, сер. Техническая кибернетика. N5, 1994. — С.143-160.
15. Гэри М, Джонсон Д. Вычислительные машины и труднорешаемые задачи. М.: Мир. 1982. — 416 с.
16. Кохов В.А., Ткаченко С.В. Программная система для исследования вычислительной сложности решения задач на графовых моделях. // Программные продукты и системы. N4, 2009. С. 137-140.
17. Незнанов А.А., Кохов В.А. Программный комплекс анализа структурного сходства систем с учетом расположения фрагментов. // Программные продукты и системы. N4, 2009. С. 147-150.