УДК 519.175.1

ИНВАРИАНТ ГРАФА НА ОСНОВЕ КОМПАКТНЫХ ПОДГРАФОВ И АЛГОРИТМ ЕГО ВЫЧИСЛЕНИЯ

Ан.В. Погребной, В.К. Погребной

Томский политехническийо университет
E-mail: avpogrebnoy@gmail.com

Для взвешенных обыкновенных графов введена оценка компактности подграфов. На основе этой оценки определён ряд инвариантов, характеризующих структуру графа с учётом неравномерности распределения значений весов по рёбрам. Основное внимание уделено понятию компактного подграфа. Предложен алгоритм выделения компактных подграфов и вычисления на их основе инвариантов. Подмечено важное свойство компактных подграфов – способность отражать эффект обособления подмножеств вершин с высокой оценкой компактности.

Ключевые слова:
Инвариант графа, взвешенный обыкновенный граф, оценка компактности подграфа, компактный подграф, эффект обособления вершин.

Key words:
Graph invariant, weighted ordinary graph, estimation of subgraph density, dense subgraph, vertex isolation effect.

Введение

Для описания и анализа структуры графовых моделей в теории графов накоплено большое разнообразие количественных показателей (дескрипторов), каждый из которых характеризует некоторое структурное свойство графа [1]. Например, хроматическое число графа, число внутренней устойчивости, вектор упорядоченных значений степеней вершин [2]. Дескрипторы (описатели), которые определяются независимо от нумерации вершин или какого-либо их наименования, т. е. являются инвариантными относительно обозначения вершин, стали называться инвариантами [2]. В частности, матрица смежности вершин графа не является инвариантом, т. к. зависит от нумерации вершин (строк и столбцов матрицы). Напротив, длина кратчайшего маршрута в графе является инвариантом, а непосредственно сам маршрут в виде последовательности вершин не является.

В общем случае можно говорить, что каждый инвариант представляет некоторую характеристику структуры абстрактного графа, т. е. графа, у которого отсутствует обозначение вершин. Эти характеристики интерпретируют различные свойства графовых моделей, разработанных для конкретных приложений.

Инварианту соответствует количественная мера в виде числа или вектора. Поэтому инварианты одного типа, полученные для разных графов, могут сравниваться между собой, определяя степень сходства или различия этих графов. В изоморфных графах инварианты совпадают, т. е. равенство графов является необходимым условием их изоморфизма. К сожалению, обратное утверждение не верно, т. к. равенство инвариантов у двух графов вовсе не гарантирует наличие у них изоморфизма. Объёмы вычислений, связанные с определением инвариантов, зависят от их типов. Некоторые из них, например степени вершин, определяются совсем просто, а такие как хроматическое число графа требуют решения отдельной сложной задачи.

Во многих практических приложениях графовые модели представляются в виде взвешенных обыкновенных графов. В таких моделях рёбрам графа приписываются некоторые значения (веса), отражающие отдельное свойство соответствующей коммуникации, например протяженность, пропускную способность и т. п. Для анализа таких графовых моделей важно иметь инвариант, который характеризует структуру графа с учётом весов рёбер.

В статье вводится инвариант данного типа и предлагается алгоритм его вычисления. Для разработки инварианта вводится новое понятие – *компактность подграфа* и даётся её количественная оценка. Предлагаемый инвариант даёт возможность связать неравномерность распределения значений весов по рёбрам графа с его структурой и оказывается полезным при решении ряда задач анализа структур графов.

Компактность подграфов и инварианты

Рассматривается связный обыкновенный граф $G=(E,U)$ с множеством вершин $E=\{e_i\}$, $i=1,2,...,n$ и множеством рёбер $U=\{u_{ij}\}$. Каждому ребру $u_{ij}=(e_i,e_j)$, связывающему две вершины e_i и e_j, ставится в соответствие вес $r_{ij}>0$. В этом случае граф G может быть представлен матрицей $R=\|r_{ij}\|$, которая совмещает матрицу смежности вершин и матрицу весов рёбер. Граф G может быть также представлен множеством инциденторов $F(e_i)$ вершин $e_i \in E$. Инцидентор $F(e_i)$ представляет совокупность вершин $e_j \in E$, $i \neq j$, которые в графе G связаны рёбрами $u_{ij}=(e_i,e_j)$, инцидентными вершине e_i. Мощность множества $F(e_i)$ в этом случае определяет степень s_i вершины e_i, т. е. $s_i=|F(e_i)|$.

Если значение веса r_{ij} ребра $(e_i\, e_j)$ ассоциировать с силой притяжения (сжатия) вершин e_i и e_j между собой, то можно говорить о компактности совокупности вершин e_i и e_j (от лат. *compactus* – «сжатый»). При этом будем считать, что чем больше значение веса r_{ij}, тем выше значение компактности.

Таким образом, в качестве количественной меры компактности подграфа, содержащего вершины e_i и e_j, может быть принято значение веса r_{ij}. Аналогичные рассуждения можно распространить на подграфы, содержащие больше двух вершин. В этом случае оценка компактности $C(E_t)$ подграфа $G_t=(E_t,U_t)$ с множеством вершин $E_t\subset E$ и множеством рёбер $U_t\subset U$ определяется в виде суммы значений весов рёбер $u_{ij}\in U_t$, связывающих вершины $e_i\in E_t$,

$$C(E_t)=\sum_{u_{ij}\in U_t} r_{ij}.$$

На основе оценок компактности подграфов можно предложить ряд инвариантов:

1) сумма весов рёбер графа (компактность графа);
2) вектор оценок компактности кратчайших маршрутов в графе;
3) вектор упорядоченных значений оценок компактности связных подграфов, содержащих заданное число вершин;
4) вектор упорядоченных значений оценок компактности подграфов, построенных для подмножеств вершин, выделенных по определённым правилам.

Инварианты первой группы вычисляются на основе оценок компактности графа в целом. Сюда можно отнести инвариант $C(E)$, равный сумме весов рёбер графа, или инвариант $C(U)=C(E)/|U|$, равный среднему значению весов рёбер. В качестве инварианта этой группы можно рассматривать также гистограмму частот значений весов рёбер в графе.

Во вторую группу отнесены инварианты, характеризующие компактность кратчайших маршрутов между всеми парами вершин или компактность подграфов, содержащих вершины этих маршрутов. В первом случае для каждой пары вершин вычисляется сумма весов рёбер кратчайшего маршрута. Значения этих сумм в векторе инварианта упорядочиваются по возрастанию или убыванию. Во втором случае элемент в векторе инварианта вычисляется как оценка компактности подграфа, содержащего вершины кратчайшего маршрута. Эта оценка суммирует веса всех рёбер, связывающих вершины маршрута в графе.

Векторы инвариантов третей группы $C^g(E^g)=\{C_t^g(E_t^g)\}$ формируются на основе оценок компактности $C_t^g(E_t^g)$ связных подграфов, содержащих E_t^g вершин. Здесь $t=1,2,...,t_g$ обозначает порядковый номер подграфа, оценка компактности которого заняла t-е место в упорядоченной по возрастанию последовательности элементов вектора $C^g(E^g)$, t_g – общее число связных подграфов, содержащих g вершин.

В отличие от третей группы векторы инвариантов четвертой группы $C(E_t)$ включают оценки компактности подграфов $G_t=(E_t,U_t)$, выделенных на определённых условиях. Например, в качестве G_t могут выступать полные подграфы графа G. В качестве множеств вершин E_t для выделения подграфов G_t могут быть приняты инциденторы $F(e_i)$ или $F(e_i)\cup e_i$.

В составе данной группы особый интерес представляют инварианты, основанные на так называемых *компактных подграфах*. Введению понятия компактного подграфа, алгоритму их выделения в графе и инварианту, полученному на основе этого понятия, посвящено последующее содержание данной статьи.

Компактным подграфом графа G будем называть связный подграф $G_i^g=(E_i^g,U_i^g)$ $i=1,2,...,n$ с множеством вершин E_i^g, $|E_i^g|=g$ и множеством рёбер U_i^g, связывающих вершины из множества E_i^g, который в множестве E_i^g содержит вершину e_i и имеет наибольшую оценку компактности G_i^g.

Для выделения компактного подграфа G_i^g в графе G нужно сформировать все связные подграфы с множеством вершин E_i^g, $e_i\in E_i^g$, $|E_i^g|=g$, получить для каждого из них оценку компактности G_i^g и выбрать подграф, соответствующий наибольшей оценке.

Инвариантом графа G на основе компактных подграфов G_i^g будем называть n-мерный вектор C^g с упорядоченными по возрастанию оценками компактности C_i^g подграфов G_i^g.

Алгоритм выделения связных подграфов

Задача выделения компактных подграфов G_i^g может быть решена путём простого перебора всех сочетаний вершин из n по g. Для каждого сочетания анализируется соответствующей подграф, и если он является связным, то для него определяется оценка компактности, а несвязные подграфы исключаются из рассмотрения. Далее в полученном множестве подграфов для каждой вершины e_i выбирается компактный подграф G_i^g, который имеет наибольшую оценку компактности и содержит вершину e_i.

Очевидным недостатком такого алгоритма является необходимость полного перебора всех сочетаний вершин и соответствующих подграфов. При анализе графа G, содержащего несколько десятков вершин, число сочетаний становится большим уже при $g\geq 4$. Поэтому важно, чтобы алгоритм мог формировать сочетания вершин, которые соответствуют только связным подграфам.

Второе требование к алгоритму обусловлено тем, что при анализе графа G как правило требуется последовательно выделять компактные подграфы для $g=2,3,4...$ и формировать соответствующие инварианты C^g. В этом случае при генерации множества сочетаний для очередного значения $(g+1)$ алгоритм не должен выполнять генерацию заново, а всего лишь продолжить её, используя множество связных подграфов, содержащих g вершин. Предлагаемый ниже алгоритм разработан с учётом данных требований.

Множество связных подграфов алгоритм формирует в виде дерева. На уровне $g=1$ в дереве располагаются вершины e_i множества E, упорядоченные по возрастанию номеров. Вершины этого множества представим как условно связанные с нулевой вершиной $(e_{i=0}^{g=0})$ на уровне $g=0$ и обозначим его в виде $Z(e_0^0)$. Вершина $e_i^{g=1}\in Z(e_0^0)$ выступает в качестве корневой вершины автономного дерева, ветви которого соответствуют всем связным подграфам, содержащим корневую вершину. Построение такого дерева заключается в формировании множеств $Z(e_i^{g+1})$, вершины которых в дереве связываются с вершинами e_i^g на предыдущем g-м уровне.

Совокупность вершин множества $Z(e_i^{g-1})$ представим в виде упорядоченной по возрастанию номеров последовательности $(e_{i_1}^g, e_{i_2}^g, ..., e_{i_k}^g, ..., e_{i_z}^g)$. Будем считать, что вершина $e_{i_k}^g \in Z(e_{i_k}^{g-1})$, для которой формируется множество $Z(e_{i_k}^g)$ вершин $e_{i_k}^{g+1}$, связанных в дереве с вершиной $e_{i_k}^g$, разбивает множество $Z(e_{i_k}^{g-1})$ на две части $Z^0(e_{i_k}^{g-1})=(e_{i_1}^g, e_{i_2}^g, ..., e_{i_k}^g)$ и $Z^1(e_{i_k}^{g-1})=(e_{i_{k+1}}^g, ..., e_{i_z}^g)$, z – число вершин в множестве $Z(e_{i_k}^{g-1})$.

В принятых обозначениях алгоритм формирует множества $Z(e_{i_k}^g)$ по следующим правилам:

$$Z(e_0^0) = E = \{e_i^1\},\ i = 1, 2, ..., n; \qquad (1)$$

$$Z(e_{i_k}^1) = F(e_{i_k}^1) \setminus Z^0(e_0^0). \qquad (2)$$

В общем случае для вершин $e_{i_k}^g$ дерева на уровне $g \geq 2$ множества $Z(e_{i_k}^g)$ определяются по выражению:

$$Z(e_{i_k}^g) = (F(e_{i_k}^g) \cup (Z^1(e_{i_k}^{g-1})) \setminus \cup_{g=0}^{g} Z^0(e_{i_k}^{g-1}). \qquad (3)$$

Так, например, для вершины $e_{i_k}^2$ на уровне g=2 имеем:

$$Z(e_{i_k}^2) = (F(e_{i_k}^2) \cup (Z^1(e_{i_k}^1)) \setminus (Z^0(e_0^0) \cup Z^0(e_{i_k}^1)). \qquad (4)$$

Выражения (1), (2) определяют подграфы, содержащие 1 и 2 вершины соответственно. Все последующие подграфы, содержащие g+1 вершину, определяются рекуррентно по выражению (3). В дереве на уровне g+1 каждой вершине $e_{i_k}^{g+1}$ соответствует подграф, содержащий все вершины ветви от $e_{i_k}^{g+1}$ до корневой вершины. Число подграфов на уровне g+1 определяется суммарным числом вершин всех множеств $Z(e_{i_k}^g)$, сформированных для вершин $e_{i_k}^g$, расположенных на уровне g дерева.

Работу алгоритма покажем на примере графа G и его матрицы R, представленных на рис. 1.

При построении дерева нет необходимости учитывать веса. Поэтому вес r_{ij}>0 рассматривается в матрице R как факт наличия ребра u_{ij}, связывающего вершины e_i и e_j в графе G. Будем также считать, что при построении дерева переход на уровень g осуществляется после формирования всех подграфов на предыдущем уровне. На рис. 2 приведено дерево, построенное по данному алгоритму для g=1,2,3.

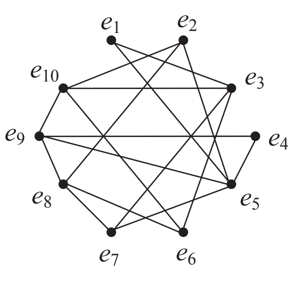
Дерево содержит 17 подграфов по 2 вершины на уровне g=2 и 41 подграф по 3 вершины на уровне g=3. Подграфы на уровне g=2 соответствуют всем рёбрам графа G. На уровнях g=2 и g=3 у каждой вершины дерева поставлены 2 числа – верхнее обозначает номер вершины e_i в графе G, а нижнее определяет компактность подграфа для соответствующей ветви дерева. Например, ветвь, выделенная жирно, соответствует подграфу с вершинами (e_2, e_8, e_7) и компактностью равной 10. Заметим, что наличие ребра в дереве не означает наличие соответствующего ребра в графе. Например, ветвь (e_2, e_5, e_{10}) содержит ребро (e_5, e_{10}), а в графе такого ребра нет. Здесь важно, что рёбра ветви (e_2, e_5, e_{10}) отражают связность подграфа.

После построения согласно выражениям (1) и (2) множеств 1-го и 2-го уровня дерева формируются множества $Z(e_{i_k}^2)$ для всех вершин $e_{i_k}^2$ 2-го уровня. Например, чтобы сформировать множество $Z(e_{8_2}^2)$ предварительно по ребру $(e_{2_2}^1, e_{8_2}^1)$ устанавливается множество $Z(e_{2_2}^1)=(e_5, e_8, e_{10})$ к которому относится вершина $e_{8_2}^2$. Далее это множество, как показано на рис. 2, разбивается на $Z^0(e_{2_2}^1)=(e_5, e_8)$ и $Z^1(ee_{2_2}^1)=(e_{10})$. Учитывая, что инцидентор $F(e_{8_2}^2)=(e_2, e_6, e_7, e_9)$, согласно (3) или в данном случае (4), имеем:

$$Z(e_{8_2}^2) = ((e_2, e_6, e_7, e_9) \cup (e_{10})) \setminus ((e_1, e_2) \cup (e_5, e_8)) =$$
$$= (e_6, e_7, e_9, e_{10}).$$

Построенное таким образом дерево удовлетворяет сформулированным ранее требованиям и содержит все связные подграфы, содержащие 3 вершины. При построении дерева алгоритм допускает формирование множеств для вершин $e_{i_k}^g$ в любой последовательности. В частности, можно без построения всего дерева автономно сформировать множество для вершины $e_{7_2}^3$, расположенной на ветви (e_2, e_8, e_7). С этой целью нужно последовательно сформировать множества $Z(e_0^0)$, $Z(e_{2_2}^1)$, $Z(e_{8_2}^2)$, $Z(e_{7_2}^3)$. В нашем примере $Z(e_{7_2}^3)$ в соответствии с (3) определяется следующим образом:

$$Z(e_{7_2}^3) =$$
$$= ((e_3, e_5, e_8) \cup (e_9, e_{10})) / ((e_1, e_2) \cup (e_5, e_8) \cup (e_6, e_7)) =$$
$$= (e_3, e_9, e_{10}).$$



	1	2	3	4	5	6	7	8	9	10
1			4		5					
2					1			6		2
3	4					2	2			6
4					7				4	
5	5	1		7			6		3	
6			2					3		5
7			2		6			4		
8		6				3	4		2	
9				4	3			2		3
10		2	6			5			3	

Рис. 1. *Пример графа G и его матрицы R*

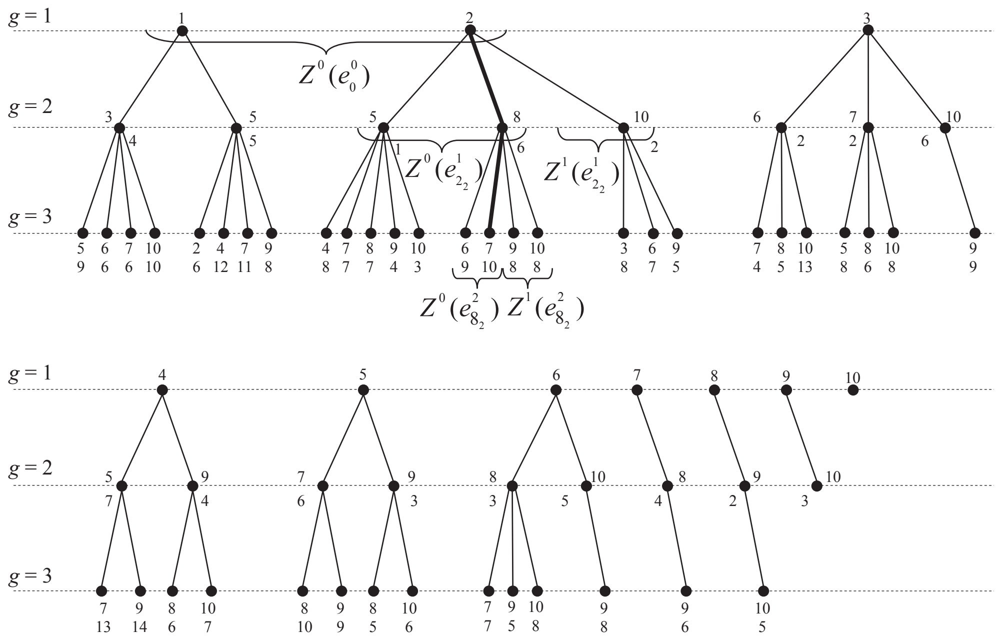


Рис. 2. *Дерево связных подграфов для g=1,2,3*

Заметим также, что упорядочение вершин в множествах $Z(e_{i_k}^g)$ по возрастанию номеров не является обязательным и введено в алгоритм всего лишь для удобства восприятия дерева.

Инварианты на основе компактных подграфов

При наличии дерева связных подграфов G_i^g с оценками компактности G_i^g выделение компактных подграфов G_i^g осуществляется последовательно для всех вершин $e_i \in E$. С этой целью среди подграфов G_i^g, содержащих вершину e_i, выбирается подграф G_i^g с максимальной оценкой компактности C_i^g. Если таких подграфов окажется несколько, то они все запоминаются. Выбор подграфов в этих случаях производится после выбора компактных подграфов для всех вершин e_i. При этом предпочтение отдаётся тому подграфу, который максимально пересекается с подграфами, выбранными для других вершин.

Результаты выбора компактных подграфов G_i^3 для рассматриваемого примера сведены в таблицу.

Столбец таблицы представляет компактный подграф G_i^3 для вершины e_i. Вершины в множестве E_i^3 графа G_i^3 указываются в последовательности, аналогичной соответствующей ветви дерева. Для вершины e_8 в дереве оказалось два компактных подграфа с оценкой C_8^3=10. Предпочтение в данном случае отдаётся подграфу с множеством вершин E_8^3=(e_2,e_8,e_7), которое полностью пересекается с E_2^3.

Таблица. *Компактные подграфы*

e_i	1	2	3	4	5	6	7	8		9	10
E_i^3	1	2	3	4	4	3	4	2	5	4	3
	5	8	6	5	5	6	5	8	7	5	6
	4	7	10	9	9	10	7	7	8	9	10
C_i^3	12	10	13	14	14	13	13	10		14	13

В нижней строке таблицы указаны значения оценок компактности. Данные оценки в упорядоченной по возрастанию последовательности образуют инвариант в виде n-мерного вектора $C^g(G)$. В нашем примере $C^3(G)$=(10,10,12,13,13,13,13,14,14,14). Нетрудно убедиться, что инвариант на основе компактных подграфов G_i^2 для дерева на рис. 2 соответствуют вектору $C^2(G)$=(4,5,5,6,6,6,6,6,7,7).

Компактные подграфы G_i^g, наряду с тем, что являются основанием для формирования инвариантов $C^g(G)$, обладают примечательным свойством, которое заключается в способности отражать степень обособления вершин в графе. Эффект обособления поясним на примере компактных подграфов G_i^3, представленных в таблице. Для этого на рис. 3 визуально покажем принадлежность вершин графа G компактным подграфам G_i^3.

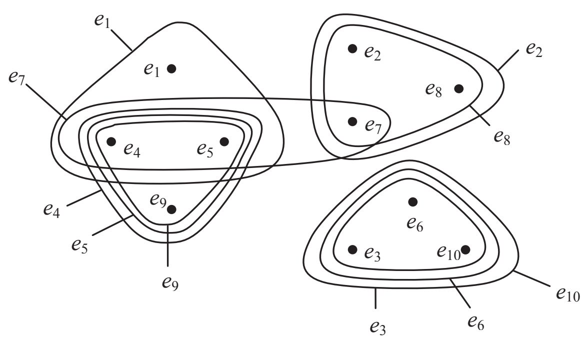


Рис. 3. *Иллюстрация эффекта обособления вершин*

Множества вершин E_i^3 на рис. 3 обведены линиями, помеченными вершинами e_i, для которых получены соответствующие компактные подграфы G_i^3.

Эффект обособления вершин легко просматривается визуально. В частности, если исключить подграф G_7^3, то вершины графа G разобьются на 3 обособленные подмножества с высокой оценкой компактности. Очевидно, что эффект обособления вершин взвешенных графов G на основе компактных подграфов G_i^g может быть использован при решении задач декомпозиции графа на подграфы. Применение компактных подграфов для решения таких задач требует отдельного исследования и в данной статье не рассматривается.

Выводы

1. Введение оценки компактности для подграфов взвешенного обыкновенного графа даёт возможность сформулировать ряд инвариантов, которые оказываются полезными при анализе структур графов с учётом весов рёбер.
2. Предложенный алгоритм формирования связных подграфов позволяет избежать полого перебора подграфов, исключая несвязные и связные дублирующие подграфы. Алгоритм может быть полезен в любых других исследованиях, где возникает потребность в полном или частичном выделении связных подграфов.
3. Введение понятия компактного подграфа и определение на этой основе инварианта в виде n-мерного вектора расширило возможность по увязке неравномерности распределения значений весов по рёбрам графа с его структурой. Кроме того, компактные подграфы выступают в роли «усилителя» эффекта обособления подмножеств вершин с высокой оценкой компактности, что может оказаться весьма полезным при решении задач декомпозиции графов.

Работа выполнена в рамках госзадания «Наука».

СПИСОК ЛИТЕРАТУРЫ

1. Дэмер М., Эммерт-Штрайб Ф., Цой Ю.Р., Вармуза К. Новый функционал информативности для анализа структуры химических графов // Известия Томского политехнического университета. – 2010. – Т. 316. – № 5. – С. 5–11.
2. Зыков А.А. Основы теории графов. – М.: Изд-во «КомКнига», 2004. – 644 с.