УДК 519.173.5:519.177

Количество простых циклов фиксированной длины в неориентированном графе. Явные формулы в случае малых длин

А. Н. Воропаев, С. Н. Перепечко

Кафедра прикладной математики и кибернетики
Петрозаводский государственный университет
Петрозаводск, 185910 Республика Карелия, Россия

Разработаны модификации алгоритма Росса и Харари для вывода формул, выражающих количество c_k простых циклов длиной k в неориентированном графе через его матрицу смежности. Рассмотрены как общий случай, так и случай двудольного графа. Алгоритмы, реализованные в системе компьютерной алгебры, позволяют выводить формулы при $k \leqslant 12$ в общем случае и при $k \leqslant 14$ в случае двудольного графа. Установлено, что при любом фиксированном $k \geqslant 8$ и затратах памяти, квадратичных относительно порядка n графа, время вычисления c_k есть величина $O(n^{[k/2]} \log n)$. Для случая двудольного графа при $k = 8$, 10, 14 установлены лучшие оценки: $O(n^3 \log^2 n)$, $O(n^4 \log^2 n)$, $O(n^6 \log^2 n)$.

Ключевые слова: алгоритмы на графах, циклы в графах, матрица смежности.

1. Введение

Задача подсчёта простых циклов возникает во многих предметных областях [1–5]. Ещё на рубеже 1940-х и 1950-х годов в социометрических исследованиях формулировалась задача подсчёта путей в орграфе. Развивавшийся в те же годы матричный подход к изучению графов способствовал выводу формул, выражающих матрицы R_k маршрутов длиной k в орграфе, не являющихся путями, через матрицу смежности. Наиболее полно результаты этого периода представлены в статье Росса и Харари [6], которые с помощью предложенного ими алгоритма воспроизвели ранее известные выражения для матриц R_3 и R_4, а также получили новые формулы — для матриц R_5 и R_6. Резкий рост размера выражений и трудоёмкости их вывода препятствовал продолжению этого ряда для бо́льших значений длины. На основе матрицы маршрутов R_{k-1} легко рассчитывается количество c_k простых циклов длиной k [1].

Выводу аналогичных матричных выражений непосредственно для величин c_k были посвящены более поздние работы [7–9]. Харари и Манвел [7] рассмотрели случаи $k = 3, 4, \ldots, 7$. Руководствуясь наглядными соображениями, для значений 3, 4 и 5 они вывели формулы, основанные на исключении из всех замкнутых маршрутов рассматриваемой длины тех, которые не являются простыми циклами. В случаях $k = 6$, 7 приведены только формы таких маршрутов (одна упущена), без соответствующих матричных выражений и коэффициентов при них в формулах для количества циклов. Авторы [8] восполнили упущение и вывели для каждой формы матричное выражение. Однако и в этой работе отсутствуют значения коэффициентов. Вывод явных формул для величин c_k оказался столь же проблематичным, что и вывод выражений для матриц R_k. Кроме того, не представлены систематические алгоритмы их вывода.

В статье [8] установлено, что сложность подсчёта циклов длиной менее 8 не превышает по порядку сложность умножения матриц размера $n \times n$, где n — количество вершин. Этот же вывод следует из работ [1, 6]. Альтернативный подход

Статья поступила в редакцию 13 декабря 2011 г.

к подсчёту простых циклов с помощью матрицы смежности представлен в статье [10]. Авторы вывели универсальную формулу для количества простых циклов произвольной фиксированной длины (за исключением случая гамильтоновых циклов, для которого приведено другое выражение). Однако сложность вычисления по этой формуле экспоненциально зависит от порядка графа.

2. Вывод формул

Алгоритм и формулы Росса и Харари [6] основаны на соотношении

$$r_{ij}^{(k)} = \sum_{S\subset X_k,\ S\neq\varnothing} (-1)^{|S|+1}|W_S|, \quad i\neq j, \tag{1}$$

где $r_{ij}^{(k)}$ — количество маршрутов длиной k из вершины i в вершину j, не являющихся путями, W_S — множество всех маршрутов длиной k из i в j, определённые вершины которых совпадают, а X_k — множество всех возможных пар номеров совпадающих вершин:

$$W_S = \{w\in W \mid \forall\{p;q\}\in S\ w_p = w_q\}, \quad X_k = \{\{p;q\}\subset 0,\dots,k \mid p<q-1\}\backslash\{\{0;k\}\}.$$

Символом W обозначено множество всех маршрутов длиной k из i в j. Количество маршрутов из множества W_S выражается с помощью элементов матрицы смежности:

$$|W_S| = \sum_{\forall\{p;q\}\in S\ \ i_p=i_q} a_{i_0i_1}a_{i_1i_2}\dots a_{i_{k-1}i_k}, \quad i_0=i, \quad i_k=j.$$

Если совпадающим (согласно S) вершинам маршрута сопоставить один и тот же индекс, то сумма преобразуется в кратную сумму (возможно, нулевой кратности), где каждый индекс пробегает диапазон $1,\dots,n$, по количеству вершин в графе. За счёт введения вспомогательных матриц, например, степеней матрицы смежности, Росс и Харари получили полностью матричную запись формул для матриц R_3, R_4, R_5 и R_6 в терминах операций «$\cdot$» (обычное умножение матриц), «$\times$» (поэлементное умножение матриц), «$^{\mathrm{T}}$» (транспонирование матрицы) и «d» (диагональная матрица с той же главной диагональю, что матрица-аргумент). Например,

$$\begin{aligned} r_{ij}^{(3)} &= \sum_{k=1}^{n} a_{ij}a_{jk}a_{kj} + \sum_{k=1}^{n} a_{ik}a_{ki}a_{ij} - a_{ij}a_{ji}a_{ij} = \\ &= \left(A\cdot d(A^2) + d(A^2)\cdot A - A\times A^{\mathrm{T}}\right)_{ij}. \end{aligned} \tag{2}$$

За исключением случая $k=3$, многие величины $|W_S|$ тождественно обращаются в нуль (в частности, все при $|S| > [(k+1)/2]\cdot[(k-1)/2]$ [6, с. 196]) или оказываются подобными, поэтому количество слагаемых в формуле (1) существенно сокращается. Например, при $k=6$ из 6475 слагаемых ($|S|\leqslant 6$) после упрощения остаётся 101.

Прямая реализация алгоритма Росса и Харари в системе компьютерной алгебры Maple (версий 7 и 12) позволила вывести формулы для матриц R_7 и R_8. Проблемой продолжения ряда явился резкий рост количества подмножеств S в формуле (1). Кроме этого, ввиду неоднозначности записи слагаемых, оставались неприведённые подобные слагаемые. Путём специального упорядочения наборов S и аналитического учёта части подобных слагаемых удалось существенно уменьшить объём перебора и получить выражения для матриц R_9 и R_{10}. В последнем случае примерно 2/3 времени заняло приведение подобных слагаемых, основанное на проверке изоморфизма соответствующих форм маршрутов.

Соотношение, аналогичное (1), имеет место и непосредственно для величин c_k. Разработанная модификация улучшенного алгоритма Росса и Харари позволила продвинуться дальше ещё на одно значение длины и получить явные выражения для c_k при $k \leqslant 12$, продолжив ряд ранее известных результатов [7–9]. Кроме того, они оказались существенно компактнее формул для матриц R_{k-1} (см. табл. 1). Алгоритм позволяет без труда учесть двудольность графа — достаточно изменить только множество X_k, исключив пары с нечётной разностью элементов. В результате получаются ещё более компактные выражения и оказывается возможным вывести формулу для количества циклов длиной 14 в двудольном графе.

Таблица 1

Количество слагаемых в формулах

k	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*14*
R_{k-1}	3	9	32	101	348	1225	4555	17475		
c_k (произвольный граф)	3	3	10	12	35	58	160	341	958	
c_k (двудольный граф)	3		7		20		59		230	1002

3. Вид слагаемых

Исходный вид каждого слагаемого в формуле (1) — кратная сумма. Преобразования Росса и Харари позволяют исключать индекс, встречающийся в паре не более чем с двумя другими индексами (см., например, (2)). Таким образом удаётся последовательно исключить все индексы в формулах для матриц R_k при $k \leqslant 6$. Величины c_k в случае произвольных графов выражаются полностью в матричном виде при $k \leqslant 7$, а в случае двудольных графов — при $k \leqslant 8$.

Количество индексов, парных с данным и отличных от него, будем называть степенью этого индекса.

Рассмотрим случай матриц R_k. Пусть y — количество индексов исходной кратной суммы степенью более двух. Цепочка индексов, например, $(i; j; k; j)$, $(i; k; i; j)$ или $(i; j; i; j)$ в выражении (2), представляет форму маршрутов. Поэтому индекс i или j (соответствующий началу или концу маршрута) участвует, по крайней мере, в одной паре соседних индексов. По этой же причине количество пар, в которых участвует любой отличный от i и j индекс, чётно. Если его степень более двух, то пар, по крайней мере, четыре. Сложив количество пар, в которых участвуют такие индексы, а также i и j, получим следующую оценку:

$$4\,y + 1 + 1 \leqslant 2\,k, \quad \text{или} \quad y \leqslant [(k-1)/2].$$

При $k \geqslant 8$ можно указать вид цепочек, для которых эта оценка достигается, причём степени всех индексов, отличных от i и j, превышают два. Например,

$$(i;\ l_1;\ l_2;\ \ldots\ ;\ l_y;\ j;\ l_y;\ l_{y-2};\ \ldots\ ; l_1;\ l_2;\ l_4;\ \ldots\ ;\ l_{y-1};\ j), \quad \text{если} \quad k \equiv 0 \pmod 4.$$

В формуле для матрицы R_7 наибольшая кратность сумм после исключения индексов равна двум.

Аналогичные рассуждения при анализе выражений для количества циклов длиной k в произвольных графах приводят к оценке $y \leqslant [k/2]$, достигаемой при $k \geqslant 8$. В случае двудольных графов удалось исследовать только частные формулы — при $k = 4, 6, \ldots, 14$. Наибольшая кратность сумм после исключения индексов равна 4, если $k = 10$, и 6, если $k = 12, 14$.

4. Упрощение формулы Хоменко и Головко

В работе [10] выведена следующая формула для количества простых циклов фиксированной длины.

$$c_k = \frac{1}{2\,k}\left|\sum_{i=0}^{n-k-1}(-1)^i \sum_{|S|=i} \operatorname{tr}(A_S^k) + \sum_{i=n-k+1}^{n-2}(-1)^i \sum_{|S|=i}\left(1+(-1)^{i-n+k}\xi_{i-n+k}\right)\operatorname{tr}(A_S^k)\right|, \tag{3}$$

$$\forall\, k \in 3, \ldots, n-1,$$

где A_S — подматрица матрицы смежности A, получаемая удалением строк и столбцов с номерами из множества $S \subset 1, \ldots, n$, $\xi_1 = n-k+1$, а остальные ξ_l удовлетворяют соотношению

$$\xi_l = -\binom{\xi_1+l-1}{l-1}\left(\frac{l-1}{l}\,\xi_1 + \frac{l-1}{\xi_1+1}\,\xi_2 + \frac{(l-1)(l-2)}{(\xi_1+1)(\xi_1+2)}\,\xi_3 + \ldots + \right.$$
$$\left. + \frac{(l-1)(l-2)\ldots 2}{(\xi_1+1)(\xi_1+2)\ldots(\xi_1+l-2)}\,\xi_{l-1}\right).$$

Два замечания позволяют упростить выражение (3). Во-первых, имеет место тождество [10, с. 388]:

$$\sum_{i=0}^{n-2}(-1)^i \sum_{S\subset V,\ |S|=i} \operatorname{tr}(A_S^k) = 0, \quad \forall\, k \in 3, \ldots, n-1.$$

С его помощью суммирование по $i \in (1, \ldots, n-2) \setminus \{n-k\}$ сводится к суммированию по $i \in n-k, \ldots, n-2$. Во-вторых, более компактно записываются коэффициенты ξ_l: $\xi_l = (-1)^{l+1}\binom{n-k+l}{n-k}$, $\quad \forall\, l \in 1, \ldots, k-2$.

В результате обоих упрощений получается выражение

$$c_k = \frac{1}{2\,k}\sum_{i=2}^{k}(-1)^{k-i}\binom{n-i}{n-k}\sum_{|S|=n-i}\operatorname{tr}(A_S^k), \quad \forall\, k \in 3, \ldots, n, \tag{4}$$

которое оказывается справедливым и при $k = n$, совпадая с формулой, приведённой в статье [10].

5. Оценка сложности формул

Под сложностью формулы в работе понимается сложность вычислений, предписанных формулой. Наибольшая кратность сумм, встречающихся в выражении для матрицы R_k при $k \geqslant 8$, равна $[(k-1)/2]$. При этом общий член суммы представляет собой произведение некоторых элементов матрицы смежности в количестве, не зависящем от порядка графа n. Сумму требуется вычислить для каждого элемента матрицы R_k, за исключением диагональных. Следовательно сложность формулы для этой матрицы при $k \geqslant 8$ есть величина $O(n^{[(k+3)/2]}\log n)$. Логарифмический множитель выражает сложность суммирования двух величин, полиномиально зависящих от n. В случае $k \leqslant 7$ выражения анализировались путём непосредственного перебора слагаемых. Сложность формулы для матрицы R_7 оценивается величиной $O(n^4 \log n)$ (встречается двойная сумма). Выражения для R_3, R_4, R_5 и R_6 записываются полностью в матричном виде. Самая трудоёмкая операция в них — умножение матриц. Следовательно при $k \leqslant 6$ сложность формул для R_k есть $O(r(n))$, где $r(n)$ обозначает сложность умножения матриц размера $n \times n$, элементы которых полиномиально зависят от n.

Случай подсчёта циклов для произвольного графа аналогичен. Наибольшая кратность сумм при $k \geqslant 8$ равна $[k/2]$, и общий член сумм такой кратности имеет тот же вид, что в случае выражений для R_k. Однако каждая сумма вычисляется единственный раз, поэтому сложность формулы для подсчёта c_k при $k \geqslant 8$ есть величина $O(n^{[k/2]} \log n)$. Выражения для c_3, c_4, …, c_7 записываются полностью в матричном виде.

Формулы для количества циклов в двудольных графах исследованы только при $k \leqslant 14$, путём перебора слагаемых. Для значений 8, 10 и 14 их сложность оказывается на порядок меньше по сравнению с выражениями в общем случае. Множитель $\log n$ заменяется на $\log^2 n$, так как в общих членах сумм наибольшей кратности встречаются произведения величин, полиномиально зависящих от n. При $k = 4, 6, 12$ сложности формул для произвольных и двудольных графов оцениваются одинаково.

В выражении (3) при произвольной фиксированной длине цикла k фигурируют почти все подмножества вершин графа. Для каждого из них вычисляется след k-й степени соответствующей подматрицы, поэтому формула имеет сложность $O(2^n r(n))$ для произвольного фиксированного k. В упрощённом варианте (4) рассматриваются всевозможные подматрицы матрицы смежности порядком не более k. Наибольшую оценку $O(n^k)$ имеет количество подматриц порядка k. Затраты на вычисление следа k-й степени подматрицы, как и её порядок, не зависят от n. Каждый след вносит в общую сумму величину, также не зависящую от n. Следовательно сложность формулы (4) имеет оценку $O(n^k \log n)$.

6. Вычислительные эксперименты

Вычисления выполнялись в системе компьютерной алгебры GAP 4.4.10 на ПК с процессором AMD Athlon 64 Processor 3500+ (2211 МГц, L1 128 КБ, L2 512 КБ), оперативной памятью DIMM DDR2 2×512 МБ, 400 МГц и операционной системой Windows XP SP 2.

Выведенные в Maple формулы были сохранены в текстовых файлах, откуда могли считываться программой, написанной на языке GAP. Для обычного умножения матриц использовались встроенные операции GAP, а вычисление кратных сумм производилось с помощью функции GAP SumX. Результаты, представленные на рис. 1, 2 практически не зависят от структуры графа, так как для любого графа выполняются одни и те же операции с его матрицей смежности.

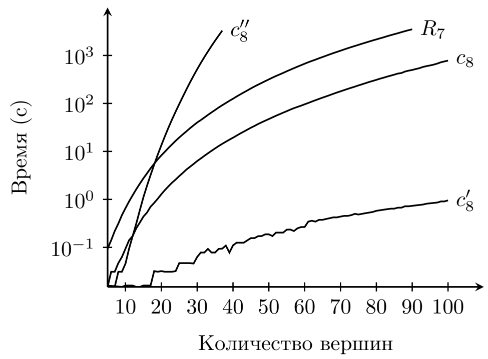

Рис. 1. Время подсчёта циклов длиной 8 по формулам в системе GAP: R_k — через матрицу маршрутов; c_k — непосредственно для произвольных графов; ${c'}_k$ — непосредственно для двудольных графов; ${c''}_k$ — по формуле (4)

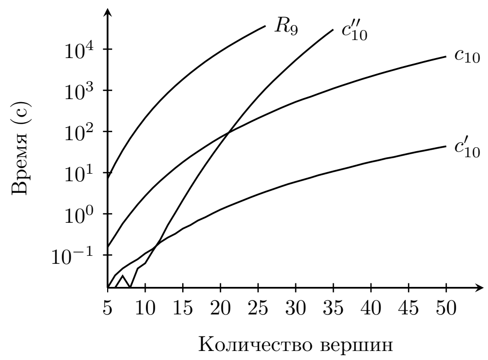

Рис. 2. Время подсчёта циклов длиной 10 по формулам в системе GAP: R_k — через матрицу маршрутов; c_k — непосредственно для произвольных графов; c'_k — непосредственно для двудольных графов; c''_k — по формуле (4)

Полученные в работе результаты применимы не только для численных расчётов, но и для символьных преобразований. В частности, для графов шахматных фигур на доске $n \times n$ были получены явные выражения зависимости количества циклов от n. Интересным продолжением работы является бинарная реализация выведенных формул, в том числе с применением параллельных вычислений.

Литература

1. *Harary F., Ross I. C.* The Number of Complete Cycles in a Communication Network // The Journal of Social Psychology. — 1954. — Vol. 40. — Pp. 329–332.
2. *Bianconi G., Capocci A.* Number of Loops of Size h in Growing Scale-Free Networks // Physical Review Letters. — 2003. — Vol. 90, No 7. — P. 078701(4).
3. Structural Properties of Planar Graphs of Urban Street Patterns / A. Cardillo, S. Scellato, V. Latora, S. Porta // Physical Review E. — 2006. — Vol. 73, No 6. — P. 066107(8).
4. *Fagiolo G.* Clustering in Complex Directed Networks // Physical Review E. — 2007. — Vol. 76. — P. 026107(8).
5. *Halford T. R., Chugg K. M.* An Algorithm for Counting Short Cycles in Bipartite Graphs // IEEE Transactions on Information Theory. — 2006. — Vol. 52, No 1. — Pp. 287–292.
6. *Ross I. C., Harary F.* On the Determination of Redundancies in Sociometric Chains // Psychometrika. — 1952. — Vol. 17, No 2. — Pp. 195–208.
7. *Harary F., Manvel B.* On the Number of Cycles in a Graph // Matematický časopis. — 1971. — Vol. 21. — Pp. 55–63.
8. *Alon N., Yuster R., Zwick U.* Finding and Counting Given Length Cycles // Algorithmica. — 1997. — Vol. 17. — Pp. 209–223.
9. *Chang Y. C., Fu H. L.* The Number of 6-cycles in a Graph // Bulletin of the Institute of Combinatorics and its Applications. — 2003. — Vol. 39.
10. *Хоменко Н. П., Головко Л. Д.* Выделение из графа его частей некоторых типов и подсчёт их количества // Украинский математический журнал. — 1972. — Т. 24, № 3. — С. 385–396. [*Khomenko N. P., Golovko L. D.* Vihdelenie iz grafa ego chasteyj nekotorihkh tipov i podschyot ikh kolichestva // Ukrainskiyj matematicheskiyj zhurnal. — 1972. — T. 24, No 3. — S. 385–396.]

UDC 519.173.5:519.177

The Number of Fixed Length Cycles in Undirected Graph Explicit Formula in Case of Small Lengths

A. N. Voropaev, S. N. Perepechko

Chair of Applied Mathematics and Cybernetics
Petrozavodsk State University
185910 Petrozavodsk, Republic of Karelia, Russia

Modifications of Ross and Harary algorithm to express the number c_k of cycles of length k in an undirected graph in terms of its adjacency matrix are developed. The general undirected graphs as well as bipartite graphs were considered. Computer algebra implementations of the algorithms enable us to construct the formulae at least for $k \leqslant 12$ in general case and for $k \leqslant 14$ in case of bipartite graph. It was shown that, for any fixed value of $k \geqslant 8$ and space complexity quadratic in order n of a graph, the time complexity of computing c_k is $O(n^{[k/2]} \log n)$. In case of bipartite graph, for $k = 8, 10, 14$ better estimations are obtained: $O(n^3 \log^2 n)$, $O(n^4 \log^2 n)$, $O(n^6 \log^2 n)$.

Key words and phrases: graph algorithms, cycles in graph, adjacency matrix.