ПРИКЛАДНАЯ ТЕОРИЯ ГРАФОВ

УДК 519.1, 519.8 DOI 10.17223/20710410/60/6

О СЛОЖНОСТИ КЛАСТЕРИЗАЦИИ ГРАФА В ЗАДАЧЕ С ОГРАНИЧЕНИЯМИ НА РАЗМЕРЫ КЛАСТЕРОВ[1]

Р. В. Балджанова*, А. В. Ильев**, В. П. Ильев*,**

*Омский государственный университет им. Ф. М. Достоевского, г. Омск, Россия,
**Институт математики им. С. Л. Соболева СО РАН, г. Омск, Россия

E-mail: baldzhanova@mail.ru, artyom_iljev@mail.ru, iljev@mail.ru

В задачах кластеризации на графах для данного графа G требуется найти ближайший к нему кластерный граф на том же множестве вершин. Граф называется кластерным, если каждая его компонента связности является полным графом. Расстояние между двумя графами понимается как число несовпадающих рёбер. Рассматривается задача кластеризации на графах, в которой размеры кластеров ограничены сверху числом s. Доказана верхняя оценка сложности кластеризации произвольного графа для случая $s = 2$. Предложен приближённый полиномиальный алгоритм решения задачи кластеризации на графах для случая $s = 3$ и доказана верхняя оценка сложности кластеризации произвольного графа для этого случая.

Ключевые слова: *граф, кластеризация, сложность кластеризации.*

ON THE COMPLEXITY OF GRAPH CLUSTERING IN THE PROBLEM WITH BOUNDED CLUSTER SIZES

R. V. Baldzhanova*, A. V. Ilev**, V. P. Il'ev*,**

*Dostoevsky Omsk State University, Omsk, Russia
**Institute of Mathematics SB RAS, Omsk, Russia

In the graph clustering problems, for a given graph G, one has to find a nearest cluster graph on the same vertex set. A graph is called cluster graph if all its connected components are complete graphs. The distance between two graphs is equal to the number of non-coincide edges. In the paper, we consider the graph clustering problem with bounded size s of clusters. The clustering complexity of a graph G is the distance from G to a nearest cluster graph. In the case of $s = 2$, we prove that the clustering complexity of an arbitrary n-vertex graph doesn't exceed $\lfloor (n-1)^2/2 \rfloor$ for $n \geqslant 2$. In the case of $s = 3$, we propose a polynomial time approximation algorithm for solving the graph clustering problem and use this algorithm to prove that clustering complexity of an arbitrary n-vertex graph doesn't exceed $(n(n-1)/2 - 3\lfloor n/3 \rfloor)$ for $n \geqslant 4$.

Keywords: *graph, clustering, clustering complexity.*

[1]Работа А. В. Ильева выполнена при поддержке гранта РНФ № 22-11-20019.

Введение

В работе изучаются задачи кластеризации, в которых заданное множество объектов требуется разбить на несколько попарно непересекающихся подмножеств — кластеров, принимая во внимание только сходство объектов друг с другом.

Задачи кластеризации вершин графов являются наиболее наглядными формализациями задач кластеризации [1]. В этих задачах вершины неориентированного графа взаимно однозначно соответствуют рассматриваемым объектам и группируются в кластеры так, чтобы минимизировать число рёбер между кластерами и число пар несмежных вершин внутри кластеров.

Будем рассматривать *обыкновенные графы*, т. е. графы без петель и кратных рёбер. Обыкновенный граф называется *кластерным*, если каждая его компонента связности является полным графом.

Существуют различные постановки задачи кластеризации вершин графа. Например, возможны варианты, при которых не накладываются никакие ограничения на кластерный граф. В других вариантах ограничивается число компонент связности или их размер. Рассматриваются также взвешенные версии задачи. Постановки и различные интерпретации задачи кластеризации вершин графов можно найти в [2–9].

1. Известные результаты

Пусть V — конечное множество. Обозначим через $\mathcal{M}(V)$ множество всех кластерных графов на множестве вершин V; $\mathcal{M}_k(V)$ — множество всех кластерных графов, имеющих ровно k компонент связности; $\mathcal{M}_{\leqslant k}(V)$ — множество всех кластерных графов, имеющих не более k компонент связности; $\mathcal{M}^{\leqslant s}(V)$ — множество всех кластерных графов, в которых размер любой компоненты связности не превосходит s.

Если $G_1 = (V, E_1)$ и $G_2 = (V, E_2)$ — обыкновенные графы на одном и том же множестве вершин V, то расстояние $d(G_1, G_2)$ между ними определяется как

$$d(G_1, G_2) = |E_1 \setminus E_2| + |E_2 \setminus E_1|.$$

Изучение кластеризации графов началось с задач без ограничений и с ограничением на количество кластеров [3, 4, 8–11]. Рассмотрим данные постановки, известные как задачи аппроксимации графов.

Задача GC. Дан граф $G = (V, E)$. Требуется найти такой кластерный граф $M^* \in \mathcal{M}(V)$, что

$$d(G, M^*) = \min_{M \in \mathcal{M}(V)} d(G, M) \stackrel{\text{def}}{=} \tau(G).$$

Задача $\mathbf{GC_k}$. Дан граф $G = (V, E)$ и целое число k, $2 \leqslant k \leqslant |V|$. Требуется найти такой кластерный граф $M^* \in \mathcal{M}_k(V)$, что

$$d(G, M^*) = \min_{M \in \mathcal{M}_k(V)} d(G, M) \stackrel{\text{def}}{=} \tau_k(G).$$

Задача $\mathbf{GC_{\leqslant k}}$. Дан граф $G = (V, E)$ и целое число k, $2 \leqslant k \leqslant |V|$. Требуется найти такой кластерный граф $M^* \in \mathcal{M}_{\leqslant k}(V)$, что

$$d(G, M^*) = \min_{M \in \mathcal{M}_{\leqslant k}(V)} d(G, M) \stackrel{\text{def}}{=} \tau_{\leqslant k}(G).$$

Будем рассматривать только задачи кластеризации с ограничением размеров компонент связности кластерного графа.

Задача $\mathbf{GC^{\leqslant s}}$. Дан граф $G=(V,E)$ и целое число s, $2\leqslant s\leqslant |V|$. Требуется найти такой кластерный граф $M^*\in\mathcal{M}^{\leqslant s}(V)$, что

$$d(G,M^*)=\min_{M\in\mathcal{M}^{\leqslant s}(V)} d(G,M)\stackrel{\text{def}}{=}\tau^{\leqslant s}(G).$$

Величины $\tau(G)$, $\tau_k(G)$, $\tau_{\leqslant k}(G)$ и $\tau^{\leqslant s}(G)$ называются *сложностью кластеризации графа* G в задачах $\mathbf{GC}$, $\mathbf{GC_k}$, $\mathbf{GC_{\leqslant k}}$ и $\mathbf{GC^{\leqslant s}}$ соответственно. Для любого n-вершинного графа G и $s\geqslant 2$ справедливо неравенство $\tau(G)\leqslant\tau^{\leqslant s+1}(G)\leqslant\tau^{\leqslant s}(G)$.

В 70-е гг. XX в. Г. Ш. Фридманом и И. Томеску были установлены точные верхние оценки сложности кластеризации графов в задачах $\mathbf{GC}$ и $\mathbf{GC_{\leqslant k}}$. В 1974 г. Г. Ш. Фридманом доказана следующая теорема:

Теорема 1 [4]. Для любого n-вершинного графа G

$$\tau(G)\leqslant\left\lfloor\frac{(n-1)^2}{4}\right\rfloor.$$

Независимо И. Томеску получил аналогичный результат для задачи $\mathbf{GC_{\leqslant k}}$:

Теорема 2 [8]. Для любого $k\geqslant 2$ и любого n-вершинного графа G

$$\tau_{\leqslant k}(G)\leqslant\left\lfloor\frac{(n-1)^2}{4}\right\rfloor.$$

Позже В. П. Ильев и Г. Ш. Фридман доказали следующую теорему:

Теорема 3 [10]. Для любого $k\geqslant 2$ и любого n-вершинного графа G при $n\geqslant 5(k-1)$

$$\tau_k(G)\leqslant\left\lfloor\frac{(n-1)^2}{4}\right\rfloor.$$

Что касается величины $\tau^{\leqslant s}(G)$, то она в общем случае не поддаётся такой же оценке. Например, для полного 10-вершинного графа $\tau^{\leqslant 2}(G)=40>\lfloor(n-1)^2/4\rfloor=20$ и $\tau^{\leqslant 3}(G)=36>\lfloor(n-1)^2/4\rfloor=20$.

2. Оценка сложности кластеризации в задаче $\mathbf{GC^{\leqslant 2}}$

Теорема 4. При $n\geqslant 2$ для любого n-вершинного графа G

$$\tau^{\leqslant 2}(G)\leqslant\frac{n(n-1)}{2}-\left\lfloor\frac{n}{2}\right\rfloor=\left\lfloor\frac{(n-1)^2}{2}\right\rfloor.$$

Доказательство. Воспользуемся индукцией по числу вершин.

Основание индукции при $n=2$. Нетрудно убедиться, что утверждение верно для 1-вершинного и 2-вершинных графов.

Пусть $n\geqslant 3$; предположим, что утверждение теоремы верно для всех графов с числом вершин, меньшим n.

Шаг индукции. Рассмотрим произвольный n-вершинный граф $G=(V,E)$. Если G — полный или пустой граф, то неравенство верно, поэтому без ограничения общности считаем, что G отличен от полного или пустого.

Выберем любую вершину $v\in V$ степени $d_G(v)\leqslant n-2$ и рассмотрим подграф $G'=(V\setminus\{v\},E')$, полученный из G удалением вершины v вместе со всеми инцидентными ей рёбрами.

По предположению индукции

$$d(G', M^*) \leqslant \frac{(n-1)(n-2)}{2} - \left\lfloor \frac{n-1}{2} \right\rfloor,$$

где M^* — оптимальное решение задачи $\mathbf{GC}^{\leqslant 2}$ на графе G', т. е. подграф графа G, представляющий собой наибольшее паросочетание в графе G' плюс (возможно) какие-то изолированные вершины.

Вернём назад вершину v. Рассмотрим подграф M графа G, полученный из M^* добавлением к паросочетанию нового независимого ребра, инцидентного вершине v (если такое ребро получится при возврате вершины v), или изолированной вершины v. Таким образом, в графе M либо на одно ребро больше, либо столько же рёбер, как и в M^*. Тогда

$$d(G, M) \leqslant d(G'M^*) + d_G(v) \text{ или } \ d(G, M) \leqslant d(G'M^*) + d_G(v) - 1.$$

В любом случае $d(G, M) \leqslant d(G'M^*) + d_G(v)$. Следовательно,

$$d(G, M) \leqslant \frac{(n-1)(n-2)}{2} - \left\lfloor \frac{n-1}{2} \right\rfloor + d_G(v) \leqslant \frac{(n-1)(n-2)}{2} - \left\lfloor \frac{n-1}{2} \right\rfloor + n - 2 =$$
$$= \frac{n^2 - 3n + 2}{2} + n - 1 - \left\lfloor \frac{n-1}{2} \right\rfloor - 1 = \frac{n^2 - n}{2} - \left\lfloor \frac{n-1}{2} \right\rfloor - 1 \leqslant \frac{n(n-1)}{2} - \left\lfloor \frac{n}{2} \right\rfloor,$$

так как $\lfloor (n-1)/2 \rfloor + 1 = \lfloor (n+1)/2 \rfloor \geqslant \lfloor n/2 \rfloor$. Поэтому

$$\tau^{\leqslant 2}(G) \leqslant d(G, M) \leqslant \frac{n(n-1)}{2} - \left\lfloor \frac{n}{2} \right\rfloor.$$

Теорема 4 доказана. ■

3. Приближённый алгоритм решения задачи $\mathbf{GC}^{\leqslant 3}$

Приведём алгоритм 1 нахождения приближённого решения в задаче $\mathbf{GC}^{\leqslant 3}$.

Алгоритм 1.

Вход: произвольный граф $G = (V, E)$.
Выход: кластерный граф $M \in \mathcal{M}^{\leqslant 3}(V)$.

1: В графе G находим максимальную K_3-упаковку, т. е. максимальное по включению множество попарно непересекающихся треугольников. Для каждого найденного треугольника удаляем из G все рёбра, которые связывают вершины треугольника с вершинами вне данного треугольника. Получаем граф $M_1 \cup G_1$, в котором подграф M_1 содержит только найденные треугольники, а G_1 — оставшиеся неудалённые рёбра графа G, не образующие треугольников.
2: В графе G_1 находим наибольшее паросочетание M_2. Множество вершин, не вошедших в паросочетание, обозначим M_3.
3: Строим кластерный граф $M \in \mathcal{M}^{\leqslant 3}(V)$: $M = M_1 \cup M_2 \cup M_3$.

Утверждение 1. Алгоритм 1 имеет трудоёмкость $O(n^3)$.

Доказательство. Рассмотрим операции, проводимые алгоритмом, и их трудоёмкости.

Обход в глубину находит цикл в графе за время $O(n + m)$, где m — количество рёбер графа. Следовательно, чтобы найти все треугольники, необходимо затратить $O(n(n + m))$ операций. Для поиска наибольшего паросочетания применяется алгоритм Эдмондса, известный как алгоритм сжатия цветков, трудоёмкость которого — $O(n^3)$ [12]. Объединение $M = M_1 \cup M_2 \cup M_3$ выполняется за один проход, т. е. за время $O(n)$.

Таким образом, итоговая трудоёмкость алгоритма 1 равна $O(n(n+m) + n^3 + n) = O(n^3)$. ■

Замечание 1. В полном графе G алгоритм 1 находит оптимальное решение.

4. Оценка сложности кластеризации в задаче $\mathbf{GC}^{\leqslant 3}$

Теорема 5. При $n \geqslant 4$ для любого n-вершинного графа G

$$\tau^{\leqslant 3}(G) \leqslant \frac{n(n-1)}{2} - 3\left\lfloor \frac{n}{3} \right\rfloor. \qquad (1)$$

Доказательство проведём в три этапа:

1) Пусть для данного n-вершинного графа G алгоритмом 1 найден кластерный граф $M \subseteq G$. В этом графе обозначим: m_3 — количество клик K_3, m_2 — количество клик K_2, m_1 — количество изолированных вершин. Заметим, что $n = 3m_3 + 2m_2 + m_1$.
2) Проверим, выполняется ли для графа M неравенство

$$\frac{n(n-1)}{2} - 3\left\lfloor \frac{n}{3} \right\rfloor - d(G, M) \geqslant 0. \qquad (2)$$

3) Если для графов G и M неравенство (2) нарушается, то докажем, что в каждом из таких случаев для графа G существует другой кластерный граф $M' \subseteq G$, такой, что $d(G, M') < d(G, M)$, причём для M' неравенство (2) выполнено.

Доказательство.

1) Заметим, что в найденном алгоритмом 1 кластерном подграфе M графа G количество удалённых алгоритмом рёбер между кластерами имеет разные ограничения в каждом из возможных случаев:

- K_3 и K_3 могут иметь между собой не более 9 соединяющих рёбер;
- K_3 и K_2 могут иметь между собой не более 6 соединяющих рёбер;
- K_3 и K_1 могут иметь между собой не более 3 соединяющих рёбер;
- K_2 и K_2 могут иметь между собой не более 2 соединяющих рёбер;
- K_2 и K_1 могут иметь между собой не более 1 соединяющего ребра;
- K_1 и K_1 не должны иметь между собой соединяющих ребер.

Ограничения на количество рёбер для кластеров K_3 определены для каждого такого случая как максимально возможное количество рёбер, соединяющих эти кластеры с другими в обыкновенном графе. Для всех прочих случаев действуют следующие рассуждения. Если бы, например, кластеры K_2 и K_2 имели больше двух соединяющих рёбер, то вместо первого из них алгоритм 1 нашёл бы клику K_3. Если бы кластеры K_2 и K_1 имели больше одного соединяющего ребра, то вместо K_2 алгоритм 1 нашёл бы клику K_3. Аналогично, если бы кластеры K_1 и K_1 были соединены ребром, то вместо них алгоритм 1 нашёл бы клику K_2.

Следовательно,

$$d(G,M) \leqslant \frac{9m_3(m_3-1)}{2} + 6m_3m_2 + 3m_3m_1 + \frac{2m_2(m_2-1)}{2} + m_2m_1 = \\ = \frac{9}{2}m_3^2 + m_2^2 + 6m_3m_2 + 3m_3m_1 + m_2m_1 - \frac{9}{2}m_3 - m_2.$$

Отметим, что

$$\frac{n(n-1)}{2} - 3\left\lfloor\frac{n}{3}\right\rfloor = \frac{(3m_3+2m_2+m_1)(3m_3+2m_2+m_1-1)}{2} - 3\left\lfloor\frac{3m_3+2m_2+m_1}{3}\right\rfloor = \\ = \frac{9m_3^2+4m_2^2+m_1^2+12m_3m_2+6m_3m_1+4m_2m_1}{2} - \frac{9m_3+2m_2+m_1}{2} - 3\left\lfloor\frac{2m_2+m_1}{3}\right\rfloor = \\ = \frac{9}{2}m_3^2 + 2m_2^2 + \frac{1}{2}m_1^2 + 6m_3m_2 + 3m_3m_1 + 2m_2m_1 - \frac{9}{2}m_3 - m_2 - \frac{1}{2}m_1 - 3\left\lfloor\frac{2m_2+m_1}{3}\right\rfloor.$$

Поэтому

$$\frac{n(n-1)}{2} - 3\left\lfloor\frac{n}{3}\right\rfloor - d(G,M) \geqslant \\ \geqslant \frac{9}{2}m_3^2 + 2m_2^2 + \frac{1}{2}m_1^2 + 6m_3m_2 + 3m_3m_1 + 2m_2m_1 - \frac{9}{2}m_3 - m_2 - \frac{1}{2}m_1 - 3\left\lfloor\frac{2m_2+m_1}{3}\right\rfloor - \\ - \left(\frac{9}{2}m_3^2 + m_2^2 + 6m_3m_2 + 3m_3m_1 + m_2m_1 - \frac{9}{2}m_3 - m_2\right) = \\ = m_2^2 + \frac{1}{2}m_1^2 + m_2m_1 - \frac{1}{2}m_1 - 3\left\lfloor\frac{2m_2+m_1}{3}\right\rfloor.$$

Таким образом, неравенство (2) будет выполнено, если

$$m_2^2 + \frac{1}{2}m_1^2 + m_2m_1 - \frac{1}{2}m_1 - 3\left\lfloor\frac{2m_2+m_1}{3}\right\rfloor \geqslant 0. \tag{3}$$

2) В свою очередь, неравенство (3) будет выполнено, если

$$m_2^2 + \frac{1}{2}m_1^2 + m_2m_1 - 2m_2 - \frac{3}{2}m_1 \geqslant 0. \tag{4}$$

Выясним, при каких $m_2, m_1 \in \mathbb{Z}_+$ неравенство (4) будет выполнено. Для этого воспользуемся методом выделения полных квадратов в левой части этого неравенства:

$$\left(m_2^2 + m_2m_1 + \frac{1}{4}m_1^2\right) - \frac{1}{4}m_1^2 + \frac{1}{2}m_1^2 - 2m_2 - \frac{3}{2}m_1 = \left(m_2 + \frac{1}{2}m_1\right)^2 + \frac{1}{4}m_1^2 - 2m_2 - \frac{3}{2}m_1 \ominus$$

Делаем замену переменных $m_2' = m_2 + m_1/2$:

$$\ominus (m_2')^2 + \frac{1}{4}m_1^2 - 2m_2' + m_1 - \frac{3}{2}m_1 = (m_2')^2 + \frac{1}{4}m_1^2 - 2m_2' - \frac{1}{2}m_1 =$$

$$= \left((m_2')^2 - 2m_2' + 1\right) - 1 + \left(\frac{1}{4}m_1^2 - \frac{1}{2}m_1 + \frac{1}{4}\right) - \frac{1}{4} = (m_2'-1)^2 + \left(\frac{1}{2}m_1 - \frac{1}{2}\right)^2 - \frac{5}{4} \ominus$$

Делаем замену переменных $m_2'' = m_2' - 1$, $m_1' = m_1/2 - 1/2$:

$$\ominus(m_2'')^2 + (m_1')^2 - \frac{5}{4} \geqslant 0.$$

Нетрудно заметить, что это неравенство задаёт внешность круга:

$$\frac{4}{5}(m_2'')^2 + \frac{4}{5}(m_1')^2 \geqslant 1.$$

При этом точка с координатами (m_1', m_2'') будет гарантированно лежать вне круга и, таким образом, удовлетворять неравенству (4), если

$$\frac{4}{5}(m_1')^2 \geqslant 1 \quad \text{или} \quad \frac{4}{5}(m_2'')^2 \geqslant 1.$$

Рассмотрим каждый из этих случаев с учётом того, что

$$\begin{cases} m_2'' = m_2 + m_1/2 - 1; \\ m_1' = m_1/2 - 1/2. \end{cases}$$

Неравенство $4/5\,(m_1')^2 \geqslant 1$ выполнено, если $(m_1/2 - 1/2)^2 \geqslant 5/4$. Наименьшее такое значение $m_1 \in \mathbb{Z}_+$ равно 4, т. е. возьмём $m_1 \geqslant 4$.

Неравенство $4/5\,(m_2'')^2 \geqslant 1$ выполнено, если $(m_2 + m_1/2 - 1)^2 \geqslant 5/4$. При $m_1 = 0$ наименьшее такое значение $m_2 \in \mathbb{Z}_+$ равно 3, т. е. возьмём $m_2 \geqslant 3$.

Таким образом, при $m_2 \geqslant 3$ либо $m_1 \geqslant 4$ неравенство (4), а значит, и неравенство (3) всегда будут выполнены.

3) Выясним, при каких $m_2 \leqslant 2$, $m_1 \leqslant 3$ неравенство (3) может нарушаться, т. е. алгоритм находит кластерный граф M, такой, что $d(G, M)$ превосходит оценку сложности кластеризации $\tau^{\leqslant 3}(G)$. Построим таблицу соответствующих значений левой части неравенства и посмотрим на его выполнимость.

m_2	m_1	Левая часть	Неравенство
0	0	0	ИСТИНА
0	1	0	ИСТИНА
0	2	1	ИСТИНА
0	3	0	ИСТИНА
1	0	1	ИСТИНА
1	1	-1	ЛОЖЬ
1	2	1	ИСТИНА
1	3	4	ИСТИНА
2	0	1	ИСТИНА
2	1	3	ИСТИНА
2	2	3	ИСТИНА
2	3	7	ИСТИНА

Рассмотрим случай $m_2 = m_1 = 1$, когда неравенство (3) может нарушаться.

Так как $n \geqslant 4$, то в этом случае в графе G должна содержаться хотя бы одна клика K_3, т. е. в найденном алгоритмом 1 кластерном графе M обязательно $m_3 \geqslant 1$.

Рассмотрим 6-вершинный подграф G' графа G, вершины которого разбиваются алгоритмом 1 на кластеры на последних шагах. Поскольку ранее обнаруженные подграфы K_3 не влияют на выполнение неравенства (3), его нарушение на единицу (как показано в таблице) может произойти именно на этапе работы с графом G', если количество удаляемых в нём рёбер между кластерами максимально возможное — 10. Напомним, что в случае нарушения этого ограничения алгоритм 1 непременно нашёл бы другие кластеры большего размера.

Однако если граф G' имеет ровно 10 рёбер, то каждая из вершин его клики $K_3 = \{v_1, v_2, v_3\}$ должна быть смежна с обеими вершинами его клики $K_2 = \{v_4, v_5\}$ и изолированной вершиной v_6. Но тогда в графе G' существуют другие кластеры: $K_3 = \{v_1, v_4, v_5\}$ и $K_3 = \{v_2, v_3, v_6\}$, для выделения которых в G' нужно удалить на два ребра меньше, чем для выделения кластеров графа M.

Таким образом, в графе G существует другой кластерный подграф M', такой, что $d(G, M) = d(G, M') + 2$; для графа M' неравенство (3) выполнено, а значит, оценка сложности кластеризации $\tau^{\leqslant 3}(G)$ в этом случае остаётся верна.

Следовательно, для любого n-вершинного графа G $(n \geqslant 4)$ существует кластерный граф M, такой, что

$$\frac{n(n-1)}{2} - 3\left\lfloor \frac{n}{3} \right\rfloor - d(G, M) \geqslant 0.$$

Поэтому

$$\tau^{\leqslant 3}(G) \leqslant d(G, M) \leqslant \frac{n(n-1)}{2} - 3\left\lfloor \frac{n}{3} \right\rfloor.$$

Теорема 5 доказана. ■

Замечание 2. Поскольку для любого $s \geqslant 3$ и любого n-вершинного графа G $\tau^{\leqslant s}(G) \leqslant \tau^{\leqslant 3}(G)$, то при $n \geqslant 4$

$$\tau^{\leqslant s}(G) \leqslant \frac{n(n-1)}{2} - 3\left\lfloor \frac{n}{3} \right\rfloor.$$

Заключение

Исследована задача кластеризации на графах, в которой размеры кластеров не превосходят числа s. Доказаны оценки сложности кластеризации для задач $\mathbf{GC}^{\leqslant 2}$ и $\mathbf{GC}^{\leqslant 3}$. Предложен приближённый полиномиальный алгоритм для задачи $\mathbf{GC}^{\leqslant 3}$, который используется при доказательстве оценки сложности кластеризации произвольного графа. Аналогичный подход может быть применён для получения оценок сложности кластеризации в отдельных случаях при $s > 3$, однако вопрос о его перспективности для доказательства оценки в случае произвольного s пока остаётся открытым.

ЛИТЕРАТУРА

1. *Schaeffer S. E.* Graph clustering // Comput. Sci. Rev. 2005. V. 1. No. 1. P. 27–64.
2. *Ляпунов А. А.* О строении и эволюции управляющих систем в связи с теорией классификации // Проблемы кибернетики. Вып. 27. М.: Наука, 1973. С. 7–18.
3. *Фридман Г. Ш.* Одна задача аппроксимации графов // Управляемые системы. 1971. Вып. 8. С. 73–75.
4. *Фридман Г. Ш.* Об одном неравенстве в задаче аппроксимации графов // Кибернетика. 1974. №3. С. 151.
5. *Bansal N., Blum A., and Chawla S.* Correlation clustering // Machine Learning. 2004. V. 56. P. 89–113.
6. *Ben-Dor A., Shamir R., and Yakhimi Z.* Clustering gene expression patterns // J. Comput. Biol. 1999. V. 6. No. 3–4. P. 281–297.

7. *Shamir R., Sharan R., and Tsur D.* Cluster graph modification problems // Discrete Appl. Math. 2004. V. 144. No. 1–2. P. 173–182.
8. *Tomescu I.* La reduction minimale d'un graphe a une reunion de cliques // Discrete Math. 1974. V. 10. No. 1–2. P. 173–179.
9. *Zahn C. T.* Approximating symmetric relations by equivalence relations // J. SIAM. 1964. V. 12. No. 4. P. 840–847.
10. *Ильев В. П., Фридман Г. Ш.* К задаче аппроксимации графами с фиксированным числом компонент // Докл. АН СССР. 1982. Т. 264. № 3. С. 533–538.
11. *Фридман Г. Ш.* Исследование одной задачи классификации на графах // Методы моделирования и обработки информации. Новосибирск: Наука, 1976. С. 147–177.
12. *Пападимитриу Х., Стайглиц К.* Комбинаторная оптимизация. Алгоритмы и сложность. М.: Мир, 1984. 510 с.

REFERENCES

1. *Schaeffer S. E.* Graph clustering. Computer Science Review, 2005, vol. 1, no. 1, pp. 27–64.
2. *Lyapunov A. A.* O stroenii i evolyutsii upravlyayushchikh sistem v svyazi s teoriey klassifikatsii [On the structure and evolution of control systems in connection with the theory of classification]. Problemy Kibernetiki, iss. 27, Moscow, Nauka Publ., 1973, pp. 7–18. (in Russian)
3. *Fridman G. Sh.* Odna zadacha approksimatsii grafov [A graph approximation problem]. Upravlyaemye Sistemy, 1971, vol. 8, pp. 73–75. (in Russian)
4. *Fridman G. Sh.* On an inequality in the graph approximation problem. Cybern. Syst. Anal., 1974, no. 10, p. 554.
5. *Bansal N., Blum A., and Chawla S.* Correlation clustering. Machine Learning, 2004, vol. 56, pp. 89–113.
6. *Ben-Dor A., Shamir R., and Yakhimi Z.* Clustering gene expression patterns. J. Comput. Biol, 1999, vol. 6, no. 3–4, pp. 281–297.
7. *Shamir R., Sharan R., and Tsur D.* Cluster graph modification problems. Discrete Appl. Math., 2004, vol. 144, no. 1–2, pp. 173–182.
8. *Tomescu I.* La reduction minimale d'un graphe a une reunion de cliques. Discrete Math., 1974, vol. 10, no. 1–2, pp. 173–179.
9. *Zahn C. T.* Approximating symmetric relations by equivalence relations. J. SIAM, 1964, vol. 12. no. 4, pp. 840–847.
10. *Il'ev V. P. and Fridman G. Sh.* On the problem of approximation by graphs with a fixed number of components. Sov. Math. Dokl., 1982, vol. 25, no. 3, pp. 666–670.
11. *Fridman G. Sh.* Issledovanie odnoy zadachi klassifikatsii na grafakh [Study of a classifying problem on graphs]. Metody Modelirovaniya i Obrabotki Informatsii, Novosibirsk, Nauka Publ., 1976, pp. 147–177. (in Russian)
12. *Papadimitriou C. H. and Steiglitz K.* Combinatorial Optimization: Algorithms and Complexity. NJ, Prentice-Hall Inc., 1982, 496 p.