

13 СЕКЦИЯ. МЕТОДОЛОГИЯ, МЕТОДЫ И ПРОГРАММНО-АЛГОРИТМИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ОБРАБОТКИ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА БОЛЬШИХ МАССИВОВ ИНФОРМАЦИИ

DOI:

МЕТОДЫ КЛАСТЕРИЗАЦИИ ОРГРАФОВ С ПОМОЩЬЮ ЛАПЛАСОВСКОЙ МАТРИЦЫ И ЕЕ СОБСТВЕННОГО ПРОЕКТОРА

Агаев Р.П.

*Институт проблем управления им. В.А. Трапезникова РАН, Россия, г. Москва
ул. Профсоюзная д.65
agaraf3@gmail.com*

Аннотация: Работа посвящена кластеризации орграфов. Актуальность таких задач связана с тем, что в классической задаче спектральной кластеризации игнорирование направлений дуг часто приводит к потере смысла связи между объектами. Предложены новые способы кластеризации орграфов, основанные на коэффициентах присоединенной матрицы лапласовской матрицы орграфа связи.

Ключевые слова: кластеризация орграфов, лапласовская матрица, остовные леса, метод Леверье-Фаддеева.

Введение

Кластеризация является одним из наиболее широко используемых методов для анализа объектов в поисковых системах, информатике, энергетических сетях, социологии, биологии и т.п. Часто кластеризация проводится для данных, на которых определена сетевая структура – обычный или ориентированный граф (орграф). В случае обычного графа спектральная кластеризация является незаменимым инструментом, поскольку для дальнейшего отображения данных в пространстве можно применить весь арсенал линейной алгебры и алгебраической теории графов.

Как обычно спектральная кластеризация проводится для неориентированного графа. Если граф является ориентированным и планируется использовать спектральный метод, то на каком-то этапе симметризация исходного орграфа неизбежна (например, как $A + A^T$ или AA^T). А если данные заданы вовсе не в виде обычного графа, то иногда применяются методы построения простого графа на множестве этих данных и в результате получается граф подобия (similarity graph). С обзором основных работ по спектральной кластеризации и ее приложениями можно ознакомиться в [1-3].

При спектральной кластеризации для построенного графа строится ненормированная лапласовская матрица L и для нее или же для одной из следующих ее нормированных матриц $L_{\text{sym}} := D^{-1/2}LD^{1/2}$ или $L_{\text{rw}} := I - D^{-1}W$ вычисляются все собственные значения. Поскольку матрицы симметричны, их спектр положителен. Далее строится матрица $U = (u_1, \dots, u_k)$ из k собственных векторов u_1, \dots, u_k для первых наименьших k собственных значений. Строки матрицы соответствуют вершинам графа в k -мерном пространстве и для них применяется алгоритм k -средних.

Различные методы симметризации [4] отношений между вершинами или методы построения подобных графов приводят к искажению исходного орграфа и потере информации. Например, в сети при кластеризации ресурсов важную роль играет направления ссылок или влияний. В таком случае игнорирование направленности сети часто приводит к нежелательному выводу. За последние 15 лет появился ряд работ, в которых при кластеризации вершин орграфов до определенного этапа сохраняется направленность исходного орграфа и используются матрицы (как диагональная матрица с элементами вектора стационарного распределения), связанные с исходным орграфом.

Тем не менее, кластеризация орграфов существенно отличается от кластеризации обычных графов. Здесь нет четкого определения кластера орграфов: в каждой предметной области, в зависимости от типов кластеров существуют свои особенности и методы кластеризации [3]. В любом случае задача кластеризации орграфа – это группировать некоторые вершины по *сходным признакам*. Существует ряд работ по кластеризации орграфов, в которых (как в случае обычного графа): 1) сперва приводится целевая функция, которая фиксирует понятие структуры кластеризации, в соответствии с выбранным определением; 2) узлам сети присваивается номер кластера, оптимизируя целевую функцию [3].

В литературе часто рассматривают два подхода кластеризации орграфов: 1) кластеризация по плотности связи (density-based) внутри кластеров, которая соответствует традиционному способу и 2)

кластеризация на основе шаблона (pattern-based) орграфов. В этом случае в один кластер попадают вершины, которые вместе создают определенный узор в общей структуре (рис. 1).

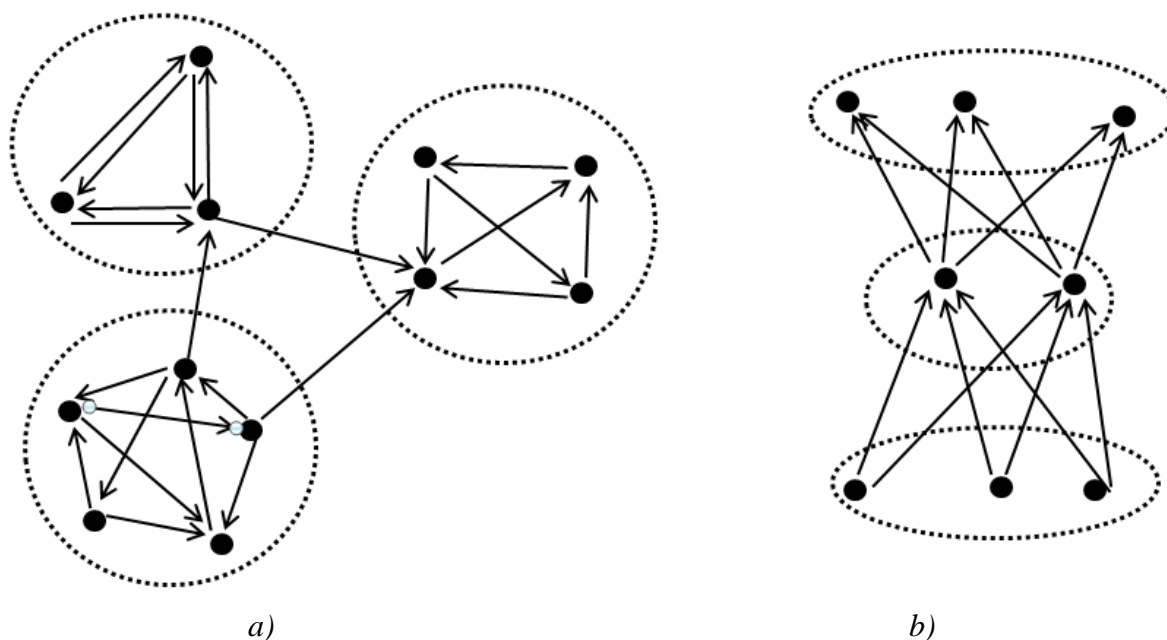


Рис. 1. Два типа кластеризации: а) по плотности; б) по шаблону (узору) кластеров

1 Базовые методы кластеризации орграфов

1.1 Метод модульных функций (modularity function)

В начале данный метод был разработан для обычных – неориентированных сетей, а потом применен для орграфов. Согласно этому методу множество объектов делится на два кластера с помощью оптимизации модульной функции [5]

$$Q = Q_1 - Q_2,$$

где Q_1 описывает суммарную долю дуг внутри кластера, Q_2 – ожидаемую долю дуг. Для орграфа в [5] предлагается следующая модульная функция

$$Q = \frac{1}{2m} \sum_{ij} \left[w_{ji} - \frac{k_i^{in} k_j^{out}}{m} \right] \delta_{c_i, c_j} = \frac{1}{2m} s^T B s;$$

где w_{ij} – вес дуги (j, i) , k_i^{in}, k_j^{out} – полустепени захода и исхода соответствующих вершин, δ_{c_i, c_j} – дельта-символ Кронекера, c_i – индикатор принадлежности i -й вершины к кластеру с номером c_i , $B = (b_{ij})$, $b_{ij} = w_{ij} - \frac{k_i^{in} k_j^{out}}{m}$, $s_i = 1$, если вершина принадлежит первому кластеру, в противном случае $s_i = -1$. Дальнейшая максимизация проводится для формы $H(Q) = \frac{1}{2m} s^T (B + B^T) s$.

1.2 Метод взвешенных разрезов

Еще один интересный способ кластеризации орграфов предложен в [6]. Результаты основаны на оценивании взвешенного сечения (разреза) орграфа и использовании диагональных матриц, элементы которых равны полустепеням исхода или же элементам вектора стационарного распределения (единственный левый собственный вектор матрицы случайного блуждания). Кластеризация проводится для элементов множества вершин орграфа, представленных в k -мерном пространстве. Также каждая вершина описывается в k -мерном пространстве и вершинам соответствуют строки матрицы $T^{-1/2} U$, со столбцами из k «наименьших» собственных векторов u_1, \dots, u_k матрицы $T^{-1/2} (2D - W - W^T) T^{-1/2}$.

Статья [7], также как [5], посвящена разделению объектов на два кластера – сообщества. Предполагается, что орграф – сильно связный и для нее построена матрица случайного блуждания. Работа представляет собой обобщением метода спектральной кластеризации и для разделения множества отмеченных и не отмеченных элементов на два кластера используется метод оптимизации. При построении базового функционала используется отношение Рэля (Rayleigh quotient) для

ориентированного графа [8]. Кластеризация неотмеченных вершин определяется знаками компонент вектора $(I - \alpha \Theta)^{-1}y$, где $\Theta = \frac{1}{2}(\Pi^{1/2}P\Pi^{-1/2} + \Pi^{-1/2}P^T\Pi^{1/2})$, Π – диагональная матрица из элементов вектора стационарного распределения, P – стохастическая матрица орграфа, y – вектор начальных отметок (состоит из 1 и -1) некоторых вершин.

Во многих задачах кластеризации орграфов предполагается, что если исходный орграф не является сильно связным и апериодичным, то аналогично задаче PageRank, используется матрица телепортации случайного блуждания (*teleporting random walk*). Хотя такой способ обеспечивает единственность левого собственного вектора для матрицы перехода, но, приводит к искажению связи между вершинами.

2 Предварительные результаты

Важно отметить, что в отличие от обычного графа, задачи кластеризации орграфов имеют широкий спектр. В одних задачах связь между вершинами внутри кластера должна быть более плотной, чем связи между кластерами. В других задачах кластеризации вершины объединяются в кластере, если они вместе находятся на пути потока маршрутов или же от них идут дуги одному множеству вершин.

Следующие обозначения взяты из [9]. Как ранее, пусть $W = (w_{ij})$ – матрица весов дуг (i, j) . Ее элемент w_{ij} равен нулю тогда и только тогда, когда из вершины i нет дуги в вершину j . Если H – подграф орграфа G , то его весом $\varepsilon(H)$ называют произведение весов всех его дуг; если в H есть вершины, но нет дуг, то $\varepsilon(H) = 1$. Если \mathcal{H} – непустое множество орграфов H , то его вес определяется как:

$$\varepsilon(\mathcal{H}) = \sum_{H \in \mathcal{H}} \varepsilon(H),$$

а если $\mathcal{H} = \emptyset$, то $\varepsilon(\mathcal{H}) = 0$.

Лапласовская матрица $L = (l_{ij})$ взвешенного орграфа $G = (V, E)$ определяется следующим образом: $l_{ij} = -w_{ji}$ и $l_{ii} = \sum_{j \neq i} w_{ij}$, где w_{ij} – вес дуги (i, j) . Исходящее дерево – это орграф, с одной отмеченной вершиной, из которой достижимы все другие вершины. В исходящем дереве для всех некорневых вершин i полустепень захода равна единице, т.е. $k_i^{in} = 1$, а если i – корень, то $k_i^{in} = 0$.

Исходящий лес – это орграф, в котором нет контуров и для всех вершин i имеет место $k_i^{in} \leq 1$.

Определение 1 [9]. Остовный исходящий лес F орграфа G называют максимальным исходящим лесом; если в G нет остовного исходящего леса с числом дуг большим, чем в F . Через \mathcal{T}_k обозначим множества всех остовных исходящих лесов орграфа с k дугами; $\mathcal{T}_k^{i \rightarrow j}$ – множество всех остовных исходящих лесов с k дугами, где j принадлежит дереву, исходящему из i .

Определение 2 [9]. Нормированной матрицей максимальных исходящих лесов орграфа называем матрицу $J_k = \sigma_k^{-1}Q_k$, где $\sigma_k = \varepsilon(\mathcal{T}_k)$, $Q_k = (q_{ji}) = (\varepsilon(\mathcal{T}_k^{i \rightarrow j}))$.

Несмотря на такое сложное определение матриц Q_k , они вычисляются очень просто, а именно согласно [9] каждая матрица рекуррентно вычисляется методом Лаверьё-Фаддеева ($Q_0 = I$):

$$(1) \quad Q_k = \frac{\text{Tr}(LQ_{k-1})}{k}I - LQ_{k-1}, \quad k = 1, \dots, n - v.$$

В [9] доказано, что все матрицы LQ_{k-1} , $k = 0, \dots, n - v - 1$ – лапласовские, а $LQ_{n-v} = 0$. Более того, все матрицы $J_k = \sigma_k^{-1}Q_k$ – стохастические.

Утверждение 1. Матрица J_{n-v-1} является невырожденной и

$$J_{n-v-1} = \left(J_{n-v} + \frac{\sigma_{n-v}}{\sigma_{n-v-1}}L \right)^{-1}.$$

Доказательство. С помощью (1) можно установить, что

$$Q_{n-v-1} = \sigma_{n-v-1}I - \sigma_{n-v-2}L + \dots + \sigma_0(-L)^{n-v} \Rightarrow J_{n-v-1} = I - \frac{\sigma_{n-v-2}}{\sigma_{n-v-1}}L + \dots + \frac{\sigma_0}{\sigma_{n-v-1}}(-L)^{n-v-1}.$$

Умножив обе части правого равенства на J_{n-v} получим

$$(2) \quad J_{n-v-1}J_{n-v} = \left(I - \frac{\sigma_{n-v-2}}{\sigma_{n-v-1}}L + \dots + \frac{\sigma_0}{\sigma_{n-v-1}}(-L)^{n-v-1} \right) J_{n-v} = J_{n-v}.$$

Также из (1) и (2) следует

$$J_{n-v} = I - \frac{\sigma_{n-v}}{\sigma_{n-v-1}} L J_{n-v-1} \Rightarrow J_{n-v-1} J_{n-v} + \frac{\sigma_{n-v}}{\sigma_{n-v-1}} L J_{n-v-1} = J_{n-v-1} \left(J_{n-v} + \frac{\sigma_{n-v}}{\sigma_{n-v-1}} L \right) = I.$$

Отметим, что из (2) непосредственно следует, как представить столбцы J_{n-v} через столбцы J_{n-v-1} .

Утверждение 2. Все диагональные элементы матрицы $J_{n-v-1} = \frac{1}{\sigma_{n-v-1}} Q_{n-v-1}$ больше нуля.

Справедливость этого утверждения следует из того, что на i -й диагонали находится сумма весов всех исходящих лесов \mathcal{F} с $n - v - 1$ дугами, где вершина i является корнем. Очевидно, что для всех i множество не пусто и $\varepsilon(\mathcal{F}) \neq 0$.

3 Кластеризация с помощью нормированных матриц исходящих лесов

3.1 Кластеризации с помощью нормированной матрицы J_{n-v-1} .

Предположим, что заданный орграф сильно связный, т.е. каждая вершина достижима из любой другой вершины или же содержит исходящее дерево. Если определим коэффициенты присоединенной матрицы для матрицы L , то для стохастической матрицы J_{n-v} выполнится $\text{rank}(J_{n-v})=1$ и все ее столбцы будут пропорциональны. Поэтому для представления вершин орграфа в пространстве мы воспользуемся только матрицей J_{n-v-1} . Действительно, согласно утверждению 1, каждый столбец J_{n-v} является линейной комбинацией столбцов матрицы J_{n-v-1} . Каждому столбцу матрицы J_{n-v-1} сопоставим числовую шкалу и предположим, что k -я ее строка представляет k -ю вершину в n -мерном пространстве.

Алгоритм 1. Задан орграф $G = (V, E)$ и число кластеров K .

1. Для заданного орграфа с $v = 1$ построить несимметричную лапласовскую матрицу $L = D - W^T$, где D – диагональная матрица полустепеней захода вершин, W – матрица смежности;
2. Методом Леверье-Фаддеева вычислить коэффициенты присоединенной матрицы B_1, \dots, B_{n-v-1} согласно формуле (1);
3. Нормировать матрицу B_{n-v-1} : $J_{n-v-1} = \frac{1}{\sigma_{n-v-1}} B_{n-v-1}$;
4. i -й строке матрицы сопоставить i -ю вершину в n -мерном пространстве;
5. Использовать алгоритм k -средних.

Предположим, что заданный орграф не является сильно связным, но содержит остовное дерево. В таком случае $v = 1$, $\text{rank}(L) = n - 1$ и $\text{rank}(J_{n-1}) = 1$, но некоторые столбцы J_{n-1} состоят только из нулей. В этом случае также можно использовать способ, описанный в предыдущем разделе.

3.2 Кластеризация с помощью нормированной матрицы J_{n-v} .

Предположим, что заданный орграф не является сильно связным, не содержит остовное дерево. Тогда $v > 1$, $\text{rank}(L) = n - v$ и $\text{rank}(J_{n-v}) > 1$. Согласно определению 2, j -й ненулевой столбец матрицы J_{n-v} соответствует усредненному весу максимальных исходящих лесов с v деревьями. А ее (i, j) элемент – усредненному весу тех лесов, в которых j является корнем дерева, а i достижима из j . При кластеризации вершин, каждой вершине будет соответствовать точка (вектор) в n -мерном пространстве. Внизу проведен алгоритм кластеризации с помощью матрицы J_{n-v} .

Алгоритм 2. Задан орграф $G = (V, E)$ и число кластеров K .

1. Для заданного орграфа с параметром $v \geq 2$ построить несимметричную лапласовскую матрицу $L = D - W^T$, где D – диагональная матрица полустепеней захода вершин, W – матрица смежности;
2. Методом Леверье-Фаддеева вычислить коэффициенты присоединенной матрицы B_1, \dots, B_{n-v} согласно формуле (1);
3. Нормировать матрицу B_{n-v} : $J_{n-v} = \frac{1}{\sigma_{n-v}} B_{n-v}$;
4. i -й строке матрицы J_{n-v} сопоставить i -ю вершину в n -мерном пространстве;
5. Использовать алгоритм k -средних.

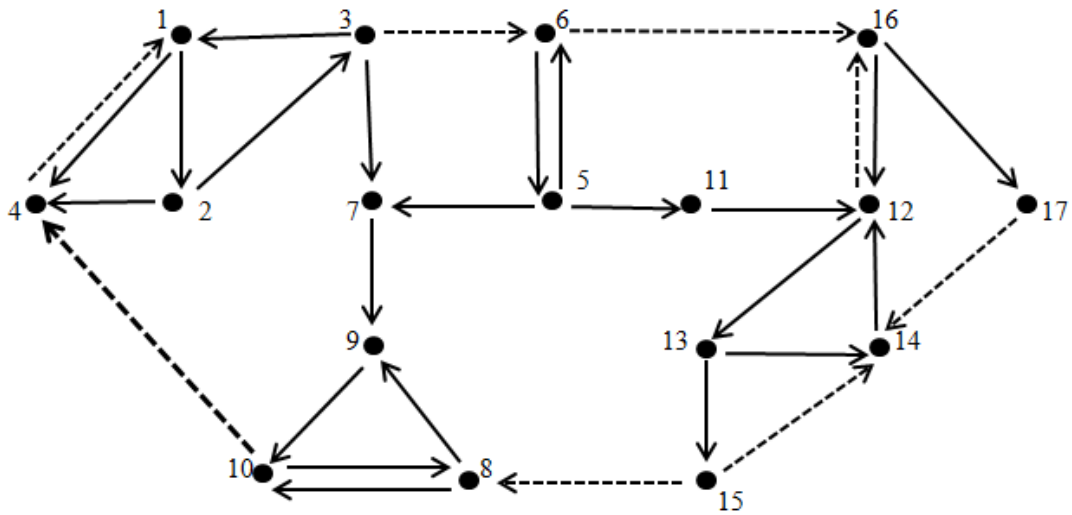


Рис.2. Орграф с тремя сильными компонентами. Пунктирные дуги дополняют орграф до сильно связного орграфа

Таблица 1. Результаты кластеризации по Алгоритму 1 для орграфа без пунктирных дуг, и по Алгоритму 2 для орграфа с пунктирными дугами.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Алгоритм 1 (2 кластера)	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2
Алгоритм 1 (3 кластера)	1	1	1	1	2	2	2	2	2	2	2	3	3	3	3	3	3
Алгоритм 1 (4 кластера)	1	1	1	1	2	2	2	2	2	2	2	3	3	3	3	4	4
Алгоритм 2 (2 кластера)	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2
Алгоритм 2 (3 кластера)	1	1	1	1	2	2	2	3	3	3	2	3	3	3	3	2	3
Алгоритм 2 (4 кластера)	1	1	1	1	2	2	2	3	3	3	2	4	4	4	4	2	4

4 Есть ли связь между методом взвешенного разреза и кластеризацией орграфа без учета направлений дуг?

Рассмотрим связь между методом кластеризации, основанной на минимизации взвешенного разреза орграфа [6, 10] и кластеризацией орграфа без учета направлений дуг.

Предположим, что вершины орграфа разбиты на кластеры $C = (C_1, \dots, C_K)$. Определим взвешенное сечение

$$(6) \quad WCut(C) = \sum_{k=1}^K \sum_{p \neq k} \frac{1}{T_k} Cut(C_k, C_p),$$

где

$$Cut(C_k, C_p) = \sum_{i, k \in C_k} \sum_{j, p \in C_p} (w_{ij} + w_{pk}),$$

T_k – вес кластера (volume) k . В качестве T_k используем сумму диагональных элементов матрицы $J_{n \times n}$ из множества C_k . Предположим, что данный метод кластеризации требует минимизировать (6). Иными словами, алгоритм считается лучшим, если для каждого кластера нормированный суммарный вес исходящих и входящих дуг минимален. Если v_1, \dots, v_K – индикаторные векторы для кластеризации $C = (C_1, \dots, C_K)$, то аналогично [6] получим:

$$WCut(C) = \sum_{k=1}^K \sum_{p \neq k} \frac{1}{T_k} Cut(C_k, C_p) = \sum_{k=1}^K \left(\frac{1}{T_k} \sum_{i \in C_k} \left(D_i^{out} + D_i^{in} - \sum_{j \in C_k} w_{ji} - \sum_{j \in C_k} w_{ij} \right) \right) =$$

$$= \sum_{k=1}^K \frac{v_{:k}^T (D^{out} + D^{in} - 2W) v_{:k}}{v_{:k}^T T v_{:k}} = \sum_{k=1}^K u_{:k}^T B u_{:k},$$

где $B = T^{-1/2} (D^{out} + D^{in} - 2W) T^{-1/2}$, $u_{:k} = T^{-1/2} v_{:k} / \sqrt{T_{kk}}$.

Далее проведут кластеризацию с помощью алгоритма k -средних, примененного для матрицы $T^{-1/2} U$, где столбцы U составлены из «наименьших» собственных векторов $\frac{1}{2}(B + B^T)$. Матрица $D^{out} + D^{in} - W - W^T$ не является лапласовской матрицей. Однако, для оценки нижней границы $WCut(C)$ часто используют ее эрмитову часть, т.е. $D^{out} + D^{in} - W - W^T$. Заметим, что полученная эрмитова часть соответствует исходному орграфу без учета направлений дуг. Данный вывод дает нам основание считать, что переход к эрмитовой части на последнем этапе кластеризации игнорирует ориентированность дуг в орграфе и такой способ подобен использованию неориентированного аналога исходного орграфа.

5 Кластеризация орграфа с помощью его конденсации

В этом разделе мы рассматриваем кластеризацию с помощью конденсации исходного орграфа. Известно, что конденсация орграфа не содержит цикла. Однако, приведенные выкладки верны для любого орграфа.

Предположим, что конденсация орграфа не является тривиальным и содержит более одной вершины с полустепенью захода, равной нулю. В отличие от исходного орграфа, конденсация орграфа всегда является ациклическим, его спектр не содержит комплексных чисел и всегда неотрицателен. Конденсация орграфа, как обычно, содержит меньшее число вершин. Проведя кластеризацию для конденсации орграфа, результат можно распространять на исходный орграф.

В качестве способа кластеризации воспользуемся Алгоритмом 2. Некоторые результаты кластеризации приведены в Таблице 2.

Особый интерес представляет разбиение вершины на два кластера. Пусть $\mathbf{s} = (s_1, \dots, s_n)^T$ – индикатор кластеризации, т.е. если вершина i принадлежит первому кластеру, то $s_i = 1$. В противном случае $s_i = -1$. Заметим, что $\delta_{ij} = \frac{1}{2}(1 - s_i s_j)$, где δ_{ij} – дельта-символ Кронекера. Как обычно, предположим, что разбиение (невзвешенное) на два класса (C_1, C_2) считается лучшим, если для данного разбиения

$$(7) \quad R = \sum_{\substack{i,k \in C_1 \\ j,p \in C_2}} (w_{ij} + w_{pk}) \rightarrow \min$$

достигает минимума.

Итак, аналогично неориентированному графу [5]

$$\begin{aligned} R &= \sum_{\substack{i,k \in C_1 \\ j,p \in C_2}} (w_{ij} + w_{pk}) = \frac{1}{2} \sum_{i,j=1}^n (1 - s_i s_j) w_{ij} = \frac{1}{2} \sum w_{ij} - \frac{1}{2} \sum s_i s_j w_{ij} = \frac{1}{2} \sum D_i^{in} - \frac{1}{2} \sum s_i s_j w_{ij} = \\ &= \frac{1}{2} \sum_{i,j=1}^n s_i s_j (\delta_{ij} D_i^{in} - w_{ij}) = \sum_{j=1}^n s_1 s_j (\delta_{1j} D_1^{in} - w_{1j}) + \dots + \sum_{j=1}^n s_n s_j (\delta_{nj} D_n^{in} - w_{nj}) = \mathbf{s}^T L \mathbf{s}. \end{aligned}$$

Заметим, что для любого орграфа (не только орграфа без цикла) $R = \mathbf{s}^T L \mathbf{s}$ достигает минимума, равного 0, при \mathbf{s} с одинаковыми компонентами. Поскольку для таких тривиальных случаев множество вершин не разбивается на два кластера, возникает вопрос: при каком векторе \mathbf{s} , $R = \mathbf{s}^T L \mathbf{s}$ будет минимальным? Данная оптимизационная задача выходит за пределы этого доклада и мы ограничимся примером, приведенным на рис. 3. Заметим, что для данного примера и всех случаев кластеризаций, полученных Алгоритмом 2, $R = \mathbf{s}^T L \mathbf{s} = 2$, что совпадает с

$$\min_{\mathbf{s} \in \pm\{1, \dots, 1\}} \mathbf{s}^T L \mathbf{s} = 2,$$

полученным перебором всех 1022 случаев кластеризаций.

Для неориентированного графа одно решение данной проблемы было предложено М. Ньюменом. Вопрос – дает ли эвристический Алгоритм 2 одно из решений оптимизационной задачи (7), представляет интерес и является предметом отдельного исследования.

На рис. 3 приведена конденсация орграфа из рис.2.

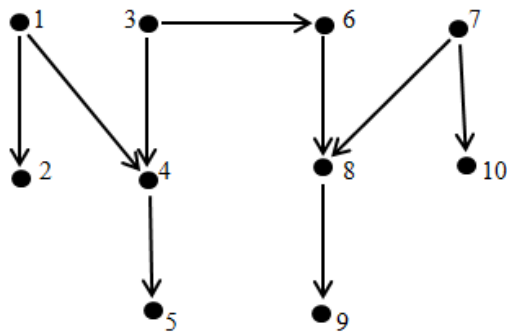


Рис.3. Конденсация (факторграф) орграфа из Рис.2 без пунктирных дуг

Таблица 2. Результат классификации для конденсации орграфа из Рис. 1 по алгоритму 2. В фигурных скобках указаны вершины сильных компонент исходного орграфа

	1 {1, 2, 3}	2 {4}	3 {5,6}	4 {7}	5 {8,9,10}	6 {11}	7 {16}	8 {12,13,14}	9 {15}	10 {17}
Два кластера	1	1	1	1	1	1	2	2	2	2
Три кластера	1	1	2	2	2	2	3	3	3	3
Четыре кластера	1	1	2	2	2	2	3	4	4	3

Заключение

В работе впервые кластеризация для орграфов проводится без введения симметричной матрицы и не используются спектральные свойства введенной дополнительной симметричной матрицы. Предложенные в настоящей работе способы кластеризации основаны на доступности и близости в орграфе и учитывают лесную и древовидную структуры орграфа.

Литература

1. Von Luxburg U. A tutorial on spectral clustering // Statistics and computing. Vol. 17. 2007, №. 4. – P. 395-416.
2. Goubko M. Ginz V. Improved spectral clustering for multi-objective controlled islanding of power grid // Energy Systems. Vol. 10. 2019, №. 1. – P. 59-94.
3. Malliaros F.D., Vazirgiannis M. Clustering and community detection in directed networks: A survey // Physics Reports. Vol. 533. 2013, №. 4. – P. 95-142.
4. Satuluri V., Parthasarathy S. Symmetrizations for clustering directed graphs // Proceedings of the 14th International Conference on Extending Database Technology. 2011. – P. 343-354.
5. Leicht E.A., Newman M.E.J. Community structure in directed networks // Physical review letters. Vol. 100. 2008, № 11. – P. 118703.
6. Meilă M., Pentney W. Clustering by weighted cuts in directed graphs // Proceedings of the 2007 SIAM international conference on data mining. – Society for Industrial and Applied Mathematics. 2007. – P. 135-144.
7. Zhou D., Huang J., Schölkopf B. Learning from labeled and unlabeled data on a directed graph // Proceedings of the 22nd international conference on Machine learning. 2005. – P. 1036-1043.
8. Chung F. Laplacians and the Cheeger inequality for directed graphs // Annals of Combinatorics. Vol. 9. 2005. – P. 1-19,
9. Agaev R.P., Chebotarev P.Y. Spanning forests of a digraph and their applications // Automation and Remote Control. Vol. 62. 2001, №. 3. – P. 443-466.
10. Shi J., Malik J. Normalized cuts and image segmentation // IEEE Transactions on pattern analysis and machine intelligence. Vol. 22. 2000, №. 8. – P. 888-905.