УДК 519.176

СПЕКТРАЛЬНЫЕ НИЖНИЕ ОЦЕНКИ ЗАТРАТ СВЯЗЫВАЮЩЕЙ СЕТИ

М.В. Губко
Институт проблем управления им. В.А. Трапезникова РАН
Россия, 117997, Москва, Профсоюзная ул., 65
E-mail: mgoubko@mail.ru

Ключевые слова: дискретная оптимизация, задача об оптимальной сети, алгебраическая теория графов, спектральные оценки

Аннотация: Рассматривается задача поиска оптимальной связывающей сети, обеспечивающей заданный набор связей между вершинами некоторого множества. Критерий оптимизации – суммарные затраты вершин сети, зависящие от входящих и исходящих из вершины потоков. Формулируется общая модель, подробно исследуется случай аддитивных функций затрат, для которых затраты вершины сети складываются из затрат, зависящих от степени вершины, и затрат, зависящих от протекающего через вершину потока. Вычисление нижней оценки затрат оптимальной сети сводится к вычислению нижних оценок отдельно для первого и второго слагаемого функции затрат. Оптимальные сети находятся для случая, когда затраты вершины зависят от ее степени. Для функции затрат, зависящей от протекающего потока предлагаются нижние оценки с использованием результатов спектральной теории графов.

1. Постановка задачи

1.1. Общая модель связывающей сети для потоковой функции затрат

Пусть задано некоторое множество $W = \{1, \ldots, n\}$ *основных вершин*.

Определение 1. Связный ориентированный граф $<W, E_R>$ с множеством вершин W и множеством дуг $E_R \subseteq W \times W$ будем называть *графом связей*.

Определение 2. Ориентированный граф H с множеством вершин V_H и множеством дуг E_H называется *сетью* над множеством основных вершин W с графом связей $<W, E_R>$, если $W \subseteq V_H$ и из того, что дуга ww' принадлежит E_R следует, что в H есть (направленный) путь из w в w'. Вершины из множества $M_H := V_H \setminus W$ будем назвать *дополнительными*. Если из контекста понятно, о какой сети идет речь, аргумент H в $V(H)$ и $E(H)$ будем опускать.

Для произвольной вершины $v \in V$ обозначим через $E^{in}(v) \subseteq E$ множество входящих в вершину дуг, а через $E^{out}(v)$ – множество исходящих дуг. Считается, что основные вершины (вершины множества W) *неспособны к коммутации*, то есть могут иметь не более одной входящей и не более одной исходящей дуги (в общем случае эти дуги могут быть инцидентны разным вершинам сети).

Определение 3. Для каждой дуги $vv' \in E(H)$ определим *функцию маршрутизации потоков* $\lambda_{vv'} : E_R \to [0,1]$, определяющую долю каждой связи $ww' \in E_R$, обеспечиваемой данной дугой, то есть долю потока, соответствующего связи ww', и текущего по дуге vv'. Если $\lambda_{vv'}(ww') = 1$, значит, дуга vv' обеспечивает связь ww' полностью. Если функция равна нулю, связь ww' данной дугой не обеспечивается, и соответствующий поток по дуге vv' не идет.

Возможность функции маршрутизации $\lambda_{vv'}(ww')$ принимать значения в интервале (0, 1) позволяет описывать разделение потока между различными маршрутами сети, связывающими вершины v и v' (т. н. «мягкую маршрутизацию»). Ниже рассматривается как модель мягкой маршрутизации, так и модель жесткой маршрутизации, в которой допустимыми значениями функции маршрутизации являются только 0 и 1, то есть запрещено *разделение потока*. Каждый раз будет поясняться, какая именно модель имеется в виду.

В сети выполняются балансовые ограничения: для любой связи ww' в начальной и конечной дуге любого маршрута из w в w' в сети H

(1) $$\lambda_{wv}(ww') = 1, \lambda_{v'w'}(ww') = 1, \text{ где } E^{out}(w) = \{wv\}, E^{in}(w') = (v'w'),$$

и в любой дополнительной вершине $m \in M_H$ выполняется равенство

(2) $$\sum_{vm \in E^{in}(m)} \lambda_{vm}(ww') = \sum_{mv \in E^{out}(m)} \lambda_{mv}(ww').$$

Затраты сети складываются из затрат ее вершин, причем основные вершины имеют нулевые затраты. Будем считать, что затраты дополнительной вершины $m \in M$ определяются только локальной структурой сети в окрестности этой вершины, а именно, входящими и исходящими потоками вершины, то есть если $E^{in}(m) = \{u_1 m, ..., u_k m\}$, $E^{out}(m) = \{mv_1, ..., mv_r\}$, то затраты дополнительной вершины m в сети H можно записать как $c(m, H) = c(\lambda_{u_1 m}(\cdot), ..., \lambda_{u_k m}(\cdot), \lambda_{mv_1}(\cdot), ..., \lambda_{mv_r}(\cdot))$. Функцию затрат такого вида будем называть *потоковой*.

Потоковые функции позволяют в максимально общем виде описывать зависимость затрат от структурных характеристик вершины m – количества k входящих и r исходящих дуг, от содержания протекающих по этим дугам потоков, а также от объема работы по коммутации потоков, которой, собственно, и занимаются дополнительные вершины сети.

Задача состоит в том, чтобы найти такую сеть H и функции маршрутизации в ней, удовлетворяющие (1) и (2) для всех связей $ww' \in E_R$, чтобы суммарные затраты всех дополнительных вершин были минимальны.

1.2. Аддитивные функции затрат

Частным случаем рассматриваемой постановки является задача поиска оптимальной *ненаправленной сети*, в которой все дуги являются двунаправленными. В такой сети все основные вершины, в силу их неспособности к коммутации, являются *висячими*. *Древовидной сетью* называется ненаправленная сеть без циклов.

Любое ребро $uv \in E_T$ древовидной сети T делит ее на две несвязных компоненты, содержащих множества основных вершин s и s' ($s \cap s' = \emptyset$, $s \cup s' = W$). Поскольку в древовидной сети разрыв любого ребра нарушает связность, все потоки между вершинами множества s и вершинами множества s' текут через ребро uv. Следовательно, потоки через любую вершину $m \in M_T$ полностью определяются количеством инцидентных ей ребер (степенью вершины) $d_T(m)$ и множествами $s_T^1(m), ..., s_T^{d_T(m)}(m) \subseteq W$ основных вершин, достижимых из вершины m через эти ребра, то есть $c(m, T) = c(s_T^1(m), ..., s_T^{d_T(m)}(m))$.

Можно показать, что задача поиска даже оптимальной древовидной сети для потоковой функции затрат эквивалентна задаче об оптимальной древовидной иерархии для т.н. *секционной функции затрат* [5]. Эта задача в общем случае имеет сверхэкспоненциальную сложность по числу n основных вершин [1], то есть рассчитывать на точные эффективные алгоритмы ее решения не приходится.

Поэтому рассматриваются частные случаи функции затрат. Например, плодотворным оказывается переход от множеств к их числовым характеристикам. Основой модели являются связи между парами вершин нижнего уровня, поэтому логично именно связь характеризовать числовыми характеристиками. В простейшем случае характеристика связи ww' одна – объем соответствующего ей информационного или материального потока $x_R(w, w')$. Когда это не приводит к путанице, будем через R также обозначать матрицу $R = (x_R(w, w'))_{w,w'\in W}$ объемов потоков между основными вершинами.

Определение 4. Потоковая функция затрат называется *аддитивной*, если она представима в виде $c(m, H) = c_1(d_H(m)) + c_2(x_H(m))$, где $d_H(m)$ – степень вершины m в сети H, $x_H(m)$ – объем потока, протекающего через вершину m в сети H, а $c_1(\cdot)$, $c_2(\cdot)$ – неотрицательные функции.

Например, пусть в древовидной сети T дополнительная вершина $m \in M$ имеет степень d, и через инцидентные ей ребра достижимы множества основных вершин $s_1, \ldots, s_d \subset W$. Тогда поток $x_T(m)$ через вершину m в сети T вычисляется как

(3) $$x_T(m) = \sum_{w,w'\in W} x_R(w,w') - \sum_{i=1}^{d} \sum_{w,w'\in s_i} x_R(w,w').$$

2. Обзор литературы

Рассматриваемая постановка близка некоторым задачам дискретной оптимизации, которые до сих пор обычно исследовались независимо. Ниже мы кратко рассмотрим их с целью подчеркнуть не только сходство с моделью связывающей сети, но и имеющиеся отличия.

2.1. Деревья и сети Штейнера

Задача построения дерева Штейнера на плоскости [2] похожа на рассматриваемую задачу тем, что в ней тоже надстраивается дерево над фиксированным множеством листьев, а множество дополнительных вершин (т.н. *точек Штейнера*) не фиксировано. Отличие, разумеется, в том, что критерий задачи Штейнера – общая длина ребер дерева, задан на ребрах, а не на вершинах, как в модели связывающей сети, и, кроме того, в связывающей сети никакой роли не играет геометрия, а лишь топология сети. В то же время известно, что задача Штейнера обычно решается в два этапа. Первый этап – выбор топологии дерева – имеет комбинаторную природу. Второй этап – выбор расположения точек Штейнера для фиксированной топологии – обычно без труда решается методами непрерывной оптимизации. Первый же этап решается фактически перебором (несколько ограниченным соображениями, специфичными именно для задачи Штейнера). Поиск оптимальной связывающей сети похож именно на первый этап решения задачи Штейнера, и нетривиальная проблема состоит в разработке высокоэффективных эвристических алгоритмов и высококачественных оценок затрат оптимальной сети. Кроме того, выше была сформулирована задача поиска оптимальной недревовидной сети. В рамках исследования задачи Штейнера также ставятся задачи поиска оптимальных сетей [10], но если в них недревовидность используется для обеспечения многосвязности сети для повышения ее надежности, в модели связывающей сети недревовидность используется для разделения потоков и гармонизации нагрузки на дополнительные вершины.

2.2. Оптимальные многоуровневые сети коммуникаций

Эта задача является обобщением задачи Штейнера на графах. В ней множество потенциальных вершин Штейнера и связей между ними ограничено графом, учитываются затраты вершин и ограниченно учитывается многоуровневость, понимаемая как возможность основывать различные части коммуникационной сети на различной элементной базе (от магистральных оптоволоконных сетей связи до локальных решений «последней мили»). Модель связывающей сети можно рассматривать как задачу построения многоуровневой сети для достаточно большого множеств потенциальных точек Штейнера, полного графа потенциальных связей и затрат вершины, зависящих не напрямую от ее типа, а от структурных свойств окружающего ее участка сети.

2.3. Задачи поиска оптимальной иерархии

Уже отмечалась связь между задачей о связывающей сети и поиском оптимальной иерархии для секционной функции затрат [1]. В частности, эти задачи эквивалентны для древовидных сетей. Аналогию можно продолжить и на другие частные случаи.

Так, в [3] исследовалась задача поиска оптимальной древовидной иерархии для *функции затрат*, *зависящей от мер*. Там для каждой основной вершины $w \in W$ задавалось число $\mu(w) > 0$ – мера вершины, и функция $c(\mu_1, \ldots, \mu_k)$ затрат вершины иерархии зависит от суммарных мер $c(\mu_1, \ldots, \mu_k)$ подмножеств основных вершин, достижимых через k дуг, входящих в вершину (дуги иерархии направлены к ее корню). Справедливо следующее утверждение:

Утверждение 1. *Задача поиска оптимальной древовидной иерархии для функции затрат, зависящей от мер, эквивалентна задаче поиска оптимальной древовидной сети в условиях потоковой функции затрат для графа потоков типа* «*звезда*» (*в котором все связи инцидентны одной вершине*).

В [5] решается задача поиска оптимальной древовидной иерархии для функции затрат вершины $c(m, H) = c_2(x_H(m))$ и графа потоков типа «цепочка» с одинаковыми потоками между соседними элементами цепи. Эта задача – частный случай поиска древовидной сети для функции затрат $c_2(x_H(m))$ и графа связей типа «симметричное кольцо».

Ниже исследуется гораздо более общая задача построения сети (в общем случае, недревовидной) для произвольного графа связей и более сложной функции затрат.

2.4. Кластеризация на графах

Как показывается далее, основным принципом построения оптимальной связывающей сети является сокращение пути в сети между сильно связанными вершинами – объединение их локальными «коммутаторами», которые, в свою очередь, соединяются друг с другом, обеспечивая менее интенсивные потоки между бóльшим числом вершин. Тот же принцип используется в задачах классификации, где для каждой пары вершин задана мера близости (роль которой в рассматриваемой модели играет объем потока) и ставится задача разбиения (возможно, иерархического) множества вершин на кластеры так, чтобы максимизировать близость вершин внутри кластеров и минимизировать близость вершин разных кластеров (см. обзор в [8]).

Отличие методов иерархической классификации от подхода настоящей работы заключается в изначально локальном подходе к оценке разбиений графа – иерархия строится в результате последовательности решений о разбиении графа на основе локального критерия, основанного на объемах межкластерных взаимодействий и на размерах кластеров (поощряется разделение на примерно равные кластеры). Иерархия обычно строится «сверху вниз», а для построения разбиений активно используются результаты спектральной теории графов [6, 7]. То, что в рамках спектральной кластеризации размер кластеров и их иерархическая структура не фиксированы, и являются предметом оптимизации, сближает эти методы с рассматриваемой моделью, однако здесь мы ставим и решаем задачу поиска иерархической кластеризации, глобально оптимальной относительно критерия, описываемого потоковой функцией.

2.5. Задачи разрезания графов

Очень близкой к задаче кластеризации на графах является задача разрезания графа, состоящая в поиске разбиения вершин графа на заданное количество групп. Это, по сути, задача кластеризации, в которой количество и размер кластеров заданы, а критерием обычно служит суммарный объем разрезаемых связей (межкластерных взаимодействий). Это NP-трудная задача целочисленного квадратичного программирования. Для решения задачи активно применяются оценки, основанные на матричных спектрах, и методы полуопределенного программирования (см. обзор в [11]). Основной технический результат нашей работы состоит в сведении задачи поиска оптимального распределения основных вершин по листьям фиксированной древовидной сети к одной из модификаций задачи разрезания графа и использовании результатов алгебраической теории графов для построения нижних оценок затрат оптимальной сети.

3. Оценка затрат оптимальной сети для аддитивной функции

Рассмотрим задачу поиска оптимальной неориентированной сети (возможно, недревовидной) для множества основных вершин W, матрицы $R = (x_R(w, w'))_{w,w'\in W}$ объемов потоков между основными вершинами в условиях жесткой маршрутизации и аддитивной потоковой функции затрат $c(m, H) = c_1(d_H(m)) + c_2(x_H(m))$.

Задача состоит в том, чтобы найти неориентированную сеть H (то есть множество дополнительных вершин M, совместно с множеством основных вершин W образующее множество V вершин сети, множество неориентированных ребер сети $E \subseteq V \times V$ и функции маршрутизации $\lambda_{vv'}(ww') \in \{0,1\}$, удовлетворяющие закону сохранения потока (1), (2) и обеспечивающую требуемые потоки $x_R(w, w')$ между всеми парами ww' основных вершин, и при этом имеющую минимальные затраты

$$C(H) := \sum\nolimits_{m\in H} c(m,H) = \sum\nolimits_{m\in H} \{c_1(d_H(m)) + c_2(x_H(m))\} = C_1(H) + C_2(H)\,.$$

Без ограничения общности матрицу R считаем симметричной. Так как граф связей $<W, E_R>$ связен, все допустимые сети также связны.

Нас будут интересовать верхние и нижние оценки затрат оптимальной сети и оценка их погрешности. Некоторые из описываемых ниже результатов относятся к задаче поиска произвольной сети, некоторые – к случаю древовидной сети, что оговаривается отдельно.

Для простоты предположим, что затраты любой вершины строго положительны. Тогда, очевидно, при поиске оптимальной сети можно ограничиться множеством сетей с априори ограниченным количеством вершин, и, следовательно, оптимальная сеть всегда существует.

Обозначим через $\Omega_i := \mathrm{Argmin}_{H\in\Omega} C_i(H)$ множество сетей, оптимальных при функции затрат вершины $c_i(\cdot)$, $i = 1, 2$. Предположим, каким-то образом найдены некоторые сети $H_i^* \in \mathrm{Argmin}_{H\in\Omega_i} C_{3-i}(H)$. Таким образом, сеть H_1^* минимизирует функцию $C_2(\cdot)$ на множестве сетей, доставляющих минимум функции $C_1(\cdot)$, а сеть H_2 – наоборот, минимизирует $C_1(\cdot)$ на множестве сетей, оптимальных при функции затрат сети $C_2(\cdot)$. Тогда выражение $\underline{C}^* := C_1(H_1^*) + C_2(H_2^*)$, очевидно, дает *нижнюю оценку* затрат $C^* := C(H^*)$ искомой оптимальной сети H^* для аддитивной функции затрат вершины $c_1(\cdot) + c_2(\cdot)$, а $C(H_1^*)$ и $C(H_2^*)$ являются *верхними оценками* затрат оптимальной сети для той же функции.

Следуя канве [4] для аналогичной задачи оптимизации сложной функции, представимой в виде суммы более простых, можно получить следующие верхние оценки относительной погрешности затрат сетей H_1^* и H_2^*, рассматриваемых как некоторые приближения к оптимальной сети H^*:

(4)
$$\varepsilon(H_1^*) := \frac{C(H_1^*) - C^*}{C^*} \; \frac{C(H_1^*) - \underline{C}^*}{\underline{C}^*} =$$
$$= \frac{C_1(H_1^*) + C_2(H_1^*) - C_1(H_1^*) - C_2(H_2^*)}{C_1(H_1^*) + C_2(H_2^*)} = \frac{C_2(H_1^*) - C_2(H_2^*)}{C_1(H_1^*) + C_2(H_2^*)}.$$

Аналогично,

(5)
$$\varepsilon(H_2^*) \; \frac{C_1(H_2^*) - C_1(H_1^*)}{C_1(H_1^*) + C_2(H_2^*)}.$$

Таким образом, представляет интерес нахождение сетей H_1^* и H_2^*. Для этого необходимо решить частные задачи поиска оптимальной сети для первого и второго слагаемого функции затрат отдельно, чему посвящены следующие разделы. Эта декомпозиция позволяет несколько упростить задачу, разбив ее на две более простых, но, увы, вторая из этих частных задач остается сложной (как мы покажем ниже, NP-трудной), сама допуская лишь приближенное решение.

Однако и эти приближенные решения можно использовать для оценки решения исходной задачи с вычислимой погрешностью.

Так, пусть как-то вычислены $\underline{C}_1$ – нижняя оценка затрат оптимальной сети для функции $c_1(\cdot)$, и $\underline{C}_2$ – нижняя оценка затрат сети для функции затрат вершины $c_2(\cdot)$. Кроме того, возьмем некоторую сеть $H_1 \in \Omega_1$, затраты которой будем использовать в качестве верхней оценки затрат $C(H_1^*)$. Заметим, что $C_1(H_1) = C_1(H_1^*)$, а $C_2(H_1) \ldots C_2(H_1^*)$.

Тогда

(6)
$$\varepsilon(H_1) = \frac{C(H_1) - C^*}{C^*} = \frac{C_1(H_1^*) + C_2(H_1) - C_1(H_1^*) - C_2(H_2^*)}{C_1(H_1^*) + C_2(H_2^*)} =$$
$$= \frac{C_2(H_1) - C_2(H_2^*)}{C_1(H_1^*) + C_2(H_2^*)} \; \frac{C_2(H_1) - \underline{C}_2}{\underline{C}_1 + \underline{C}_2} \; \frac{C_2(H_1) - \underline{C}_2}{\underline{C}_1 + \underline{C}_2}.$$

Эта оценка уже не требует для своего вычисления точного знания затрат сетей H_1^* и H_2^* – достаточно знания нижних оценок $\underline{C}_1$, $\underline{C}_2$ и верхней оценки $C_2(H_1)$. Правда, вычисление этой верхней оценки требует умения проверять принадлежность произвольной допустимой сети множеству Ω_1. Это, в свою очередь, предполагает знание величины $C_1(H_1^*)$. Ниже показывается, что задача характеризации множества Ω_1 довольно проста. Заметим, что $C_1(H_1^*)$ может использоваться в качестве точной оценки $\underline{C}_1$.

Если $\underline{C}_1 = C_1(H_1^*)$ то следующее выражение позволяет оценить относительную погрешность $\underline{C} := \underline{C}_1 + \underline{C}_2$, нижней оценки затрат сети для аддитивной функции:

$$\underline{\varepsilon} := \frac{C^* - \underline{C}}{C^*} \; \frac{C(H_1) - \underline{C}}{\underline{C}} = \frac{C_1(H_1) + C_2(H_1) - \underline{C}_1 - \underline{C}_2}{\underline{C}_1 + \underline{C}_2} = \frac{C_2(H_1) - \underline{C}_2}{\underline{C}_1 + \underline{C}_2}.$$

Заметим, что $\underline{\varepsilon} = \varepsilon(H_1)$.

Таким образом, для поиска приближенно оптимальной сети H_1 и вычисления относительной погрешности ее затрат необходимо уметь решать задачу поиска оптимальной сети для функции затрат $c_1(d_H(m))$ вершины m, зависящей только от количества $d_H(m)$ ребер, инцидентных вершине m, а также уметь решать оценочную задачу для функции затрат вершины $c_2(x_H(m))$, зависящей от объема суммарного потока $x_H(m)$ через вершину m.

4. Оптимальные сети для функций затрат $c_1(\cdot)$ и $c_2(\cdot)$

Рассмотрим функцию затрат вершины $c(m, H) = c_1(d_H(m))$, зависящей от количества $d_H(m)$ ребер, инцидентных вершине m в сети H.

Утверждение 2. *Если функция $c_1(\cdot)$ монотонна, существует оптимальная древовидная сеть. Если $c_1(\cdot)$ строго монотонна, множество Ω_1 состоит только из древовидных сетей.*

Следствие 1. *При строго монотонной функции $c_1(\cdot)$ множество Ω_1 состоит только из древовидных сетей.*

Таким образом, при строго монотонной функции $c_1(\cdot)$ сеть $H_1 \in \operatorname{Argmin}_{H \in \Omega_1} C_2(H)$ достаточно искать среди деревьев.

Лемма 1. *Пусть для* $d, d' = 3,4,...$ $c_1(d) + c_1(d') > c_1(d + d' - 2)$. *Тогда множество* $\Omega_1 := \operatorname{Argmin}_{H \in \Omega} C_1(H)$ *включает только сеть с одной дополнительной вершиной (такую сеть будем называть «звездой»).*

Следствие 2. *Лемма* 1 *верна как для вогнутых функций* $c_1(\cdot)$, *так и для субаддитивных функций* (*для которых для любых* d, d' $c_1(d) + c_1(d') > c_1(d + d')$.

Заметим, что лемма 1 верна даже для немонотонных функций затрат. В условиях этой леммы задача решена, и искомая сеть H_1 найдена, $\underline{C}_1 = C(H_1) = c_1(n)$, а для оценки погрешности $\varepsilon(H_1)$ ее затрат как верхней оценки затрат оптимальной сети для аддитивной функции необходимо найти $\underline{C}_2$, чему посвящен следующий раздел. Если же функция затрат $c_1(\cdot)$ не удовлетворяет условиям леммы, оптимальными могут быть и сети с более чем одной вершиной.

Пользуясь эквивалентностью задачи поиска оптимальной древовидной сети и исследованной в [3] задачи поиска оптимальной древовидной иерархии для однородной (степени ноль) функции затрат, можно доказать следующий результат.

Утверждение 3. *Если $H \in \Omega$, то*

$$C_1(H) \geq (n-2) \min_{d=3,\dots,n} \frac{c_1(d)}{d-2}.$$

Если при этом $n - 2$ нацело делится на $d^* - 2$, где

$$d^* = \arg\min_{d=3,\dots,n} \frac{c_1(d)}{d-2},$$

то существует d^*-однородное дерево с n листьями, и множество Ω_1 состоит из всех таких деревьев.

Для выпуклой $c_1(\cdot)$ можно показать, что множество Ω_1 для любого n состоит либо из деревьев с вершинами степеней d^* и $d^* - 1$, либо из деревьев с вершинами степеней d^* и $d^* + 1$. При этом количество вершин каждой степени характеризуется с помощью довольно простых соображений.

Рассмотрим теперь функцию затрат вершины $c(m, H) = c_2(x_H(m))$ вершины m, зависящей от суммарного объема потока $x_H(m)$, протекающего через вершину.

Утверждение 4. *Если функция $c_2(x)$ положительна при $x \geq 0$ и строго монотонна, то минимум $C_2(H)$ на Ω достигается на некоторой двухуровневой сети, имеющей вид* *клики из $p \in \{1, \ldots, n\}$ дополнительных вершин, каждой из которых инцидентно $k_m \geq 1$ висячих вершин, $m = 1, \ldots, p$.*

Следствие 3. *Пусть вдобавок к условиям утверждения 4 для любых x_1, $x_2 \geq 0$ $c_2(x_1) + c_2(x_2) \geq c_2(x_1 + x_2)$. Тогда минимум $C_2(H)$ на множестве всех неориентированных сетей достигается на «звезде» с единственной дополнительной вершиной.*

Таким образом, оптимальная сеть найдена для случая субаддитивной функции затрат, зависящей от объема протекающего через вершину потока. Для произвольной монотонной функции задача поиска оптимальной сети сведена к нахождению оптимального количества дополнительных вершин и оптимального распределения основных вершин по этим дополнительным вершинам.

5. Нижние оценки затрат сети для функции $c_2(\cdot)$

Обозначим через I_n единичную $n\times n$ матрицу, через E_n – $n\times n$ матрицу из единиц, через 1_n – n-мерный вектор-столбец, все компоненты которого равны единице. Для квадратной $n\times n$ симметричной матрицы A через $\lambda_i(A)$, $i = 1, \ldots, n$, обозначим ее *собственные числа*, перечисляемые по убыванию.

Результат, который будет использоваться далее для получения нижних оценок, был сформулирован в [7] следующим образом (с точностью до обозначений):

Утверждение 5 [7]. *Пусть заданы две квадратные симметричные действительнозначные матрицы: R размера $n\times n$ и B размера $p\times p$, причем $p \leq n$. Тогда*

(7) $$\max\{trZ^T RZB : Z^T Z = I_p\} = \max\{\sum_{i=1}^{p} \lambda_i(B)\cdot\lambda_{\phi(i)}(R) : \phi\{1,\ldots,p\} \to 1,\ldots,n, \phi \text{ инъективно}\},$$

где Z – произвольные $n\times p$ действительнозначные матрицы с взаимно ортогональными столбцами (что эквивалентно выполнению $Z^TZ = I_p$). При этом если максимум в правой части достигается на некоторой инъекции ϕ, то максимум в левой части достигается при $Z = U_1U_2^T$, где $p\times p$ матрица U_2 составлена из нормированных на единицу собственных векторов матрицы B, соответствующих собственным числам $\lambda_1(B), \ldots, \lambda_1(B)$, а $n\times p$ матрица U_1 составлена из нормированных на единицу собственных векторов матрицы R, соответствующих собственным числам $\lambda_{\phi(1)}(R),\ldots,\lambda_{\phi(p)}(R)$.

Следствие 4 [7]. *Если вдобавок к условиям утверждения 5 матрица B положительно полуопределена, то*

(8) $$\max\{trZ^T RZB : Z^T Z = I_p\} = \sum_{i=1}^{p} \lambda_i(B)\cdot\lambda_i(R).$$

Аналогичные результаты справедливы и для задачи минимизации с учетом обратного порядка перечисления собственных чисел матрицы R – от наименьшего собственного числа к бо́льшим.

Рассмотрим задачу минимизации затрат клики дополнительных вершин выбором распределения основных вершин по дополнительным в условиях функции затрат $c_2(x_H(m))$, зависящей только от суммарной величины потока через вершину.

Пусть фиксировано количество дополнительных вершин p и количество k_m основных вершин, инцидентных дополнительной вершине m (причем $\sum_{m=1}^{p} k_m = n$). Введем вектор-столбец $k := (k_1,\ldots,k_p)$ и матрицу $K = diag(k)$. Также обозначим через $s_1, \ldots, s_p$ некоторое допустимое разбиение множества основных вершин W.

Введем $n\times p$ матрицу Y, элемент которой $y_{wm} = 1$ если вершина $w \in W$ инцидентна вершине $m \in M = \{1, \ldots, p\}$ ($w \in s_m$), и нулю в противном случае. В [7] показывается, что Y тогда и только тогда описывает некоторое распределение основных вершин по дополнительным, когда:

1) $Y^T 1_n = k$,
2) $Y \cdot 1_p = 1_n$,
3) $Y \geq 0$,
4) $Y^T Y = K$.

Заметим, что ограничение целочисленности y_{wm} в явном виде отсутствует.

Обозначим через Ξ множество матриц, соответствующих разбиениям множества W на p подмножеств. Любой матрице $Y \in \Xi$ соответствует некоторое разбиение $s_1, \ldots, s_p$ n основных вершин по p дополнительным вершинам, а y_m – m-й столбец матрицы Y – является индикатором множества s_m.

В дополнительную вершину m втекают потоки ww', $w, w' \in s_m$, а также потоки ww' и $w'w$, $w \in s$, $w' \notin s$. Поэтому суммарный поток через вершину m записывается как

(9) $$x_H(m) = 2 \cdot 1_n^T R y_m - y_m^T R y_m = y_m^T \tilde{R} y_m,$$

где $\tilde{R} := 2diag(R \cdot 1_n) - R$ –несколько модифицированная матрица Лагранжа матрицы R. Заметим, что в силу того, что $y_{wm} \in \{0, 1\}$, $y_{wm} = y_{wm}^2$. Для получения (9) делается замена y_{wm} на y_{wm}^2 : $1_n^T R y_m = \sum_{i=1}^n (1_n^T R)_i \, y_{im} = \sum_{i=1}^n (1_n^T R)_i \, y_{im}^2 = y_m^T diag(R \cdot 1_n) y_m$.

Итак, задача состоит в том, чтобы найти

$$C_2(H_2) = \min_{Y \in \Xi} \sum_{m=1}^{p} c_2(y_m^T \tilde{R} y_m)$$

по всем $n \times p$ матрицам $Y = (y_1, \ldots, y_p)$ из Ξ.

Базовой версии спектральной релаксации [6], которая и будет использована ниже, соответствует минимизация по множеству матриц, удовлетворяющих лишь ограничению $Y^T Y = K$.

Поскольку известно, что для субаддитивной (в частности, для вогнутой) функции затрат оптимальна сеть со структурой «звезды», рассмотрим задачу для выпуклой функции $c_2(\cdot)$. Чтобы свести задачу к минимизации квадратичной формы, как в утверждении 5, линеаризуем функцию затрат. Оценим затраты каждой дополнительной вершины m снизу, заменив их касательной с наклоном α_m. Так как для любых чисел $\alpha, \rho \ldots 0$ $c_2(x) \ldots c_2(d_2(\alpha)) + \alpha(x - d_2(\alpha))$, где $d_2(\alpha) := [c_{2'}]^{-1}(\alpha)$, верны неравенства

(10) $$C_2(H_2) \ldots \sup_{\alpha_1,\ldots,\alpha_p \ldots 0} \min_{Y \in \Xi} \sum_{m=1}^{p} \{c_2(d_2(\alpha_m)) - \alpha_m d_2(\alpha_m) + \alpha_m y_m^T \tilde{R} y_m\} =$$
$$= \sup_{\alpha_1,\ldots,\alpha_p \ldots 0} \{\sum_{m=1}^{p} [c_2(d_2(\alpha_m)) - \alpha_m d_2(\alpha_m)] + \min_{Y \in \Xi} \sum_{m=1}^{p} \alpha_m y_m^T \tilde{R} y_m\} =$$
$$= \sup_{\alpha_1,\ldots\alpha_p \ldots 0} \{\sum_{m=1}^{p} [c_2(d_2(\alpha_m)) - \alpha_m d_2(\alpha_m)] + \min_{Y \in \Xi} trY^T \tilde{R} Y A\},$$

где $A := diag(\alpha_1, \ldots, \alpha_p)$. Также обозначим через $\alpha := (\alpha_1, \ldots, \alpha_p)$ соответствующий вектор-столбец. Заметим, что если Z – такая $n \times p$ матрица, что $Z^{\mathrm{T}} Z = 1_p$ и $Y = ZK^{1/2}$, то $Y^{\mathrm{T}} Y = K^{1/2} Z^{\mathrm{T}} Z K^{1/2}$. Поэтому

(11) $$\min_{Y \in \Xi} trY^T \tilde{R} Y A \ldots \min_{Y: Y^T Y = K} trY^T \tilde{R} Y A = \min_{Z: Z^T Z = I_p} trK^{1/2} Z^T \tilde{R} Z K^{1/2} A = \min_{Z: Z^T Z = I_p} trZ^T \tilde{R} Z A K.$$

Последнее равенство верно в силу циклического свойства следа матрицы и того, что $K^{1/2} A K^{1/2} = AK = KA$ по причине диагональности A. Очевидно, матрица $AK = diag(\alpha_1 k_1, \ldots, \alpha_p k_p)$ положительно определена, и, следовательно, мы попадаем под условие следствия 4. Поэтому

(12) $$\min_{Y\in\Xi} trY^T\tilde{R}YA \ldots \min_{Z:Z^TZ=I_p} trZ^T\tilde{R}ZAK = \sum_{m=1}^{p}\lambda_m(AK)\cdot\lambda_{n-m+1}(\tilde{R}) = \sum_{m=1}^{p}\alpha_m k_m \lambda_{n-m+1}(\tilde{R}),$$

где дополнительные вершины упорядочены убыванию $\alpha_m k_m$, и

(13) $$C_2(H_2) \ldots \sup_{\alpha_1,\ldots\alpha_p\ldots 0}\sum_{m=1}^{p}[c_2(d_2(\alpha_m)) - \alpha_m d_2(\alpha_m) + \alpha_m k_m \lambda_{n-m+1}(\tilde{R})].$$

При этом матрица, доставляющая минимум (11), вычисляется как $Y^* = U_1 U_2^T K^{1/2}$, где U_1 – $n\times p$ матрица, столбцы которой – нормированные на единицу (в евклидовой метрике) собственные вектора матрицы $\tilde{R}$, соответствующие p ее наименьшим собственным числам, упорядоченным по возрастанию, а U_2 – матрица, столбцы которой – собственные вектора матрицы AK, нормированные на единицу и упорядоченные по убыванию собственных чисел матрицы AK: $\alpha_m k_m$, где $m = 1, \ldots, p$.

Тем самым доказано

Утверждение 6. *Если функция затрат* $c_2(\cdot)$ *выпукла, а двухуровневая иерархия* H *содержит* p *дополнительных вершин, имеющих по* $k_m \geq 0$ *инцидентных висячих вершин, то верна нижняя оценка* (13).

Оно дает решение $\underline{C}_2$ оценочной задачи для выпуклой функции $c_2(\cdot)$ (напомним, что для вогнутой $c_2(\cdot)$ оптимальна «звезда» с затратами $\underline{C}_2 = c_2(\sum_{w,w'\in W} x_R(w,w'))$.

Рассмотрим задачу поиска сети H_1, оптимальной для функции затрат $c_2(\cdot)$ на множестве Ω_1, состоящем из всех деревьев с заданным числом вершин q фиксированных степеней $r_1, \ldots, r_q$, определяемых утверждением 3 и замечанием после него.

В случае древовидной сети с фиксированными q и $r_1, \ldots, r_q$ для этого необходимо определить:

1) набор ребер между дополнительными вершинами множества $M = \{1, \ldots, q\}$, то есть множество ребер $E_M \subseteq M \times M$,
2) количество $p \leq q$ дополнительных (внутренних) вершин, имеющих инцидентные основные (висячие) вершины,
3) количество $k_m < r_m$ основных вершин, подчиненных каждой дополнительной вершине $m \in M$ (без ограничения общности считаем, что $k_m = 0$ для всех $m > p$),
4) разбиение $W_1, \ldots, W_p$ ($\cup_{m=1,..p} W_m = W, W_m \cap W_{m'} = \varnothing$, $|W_m| = k_m$) основных вершин по инцидентным им дополнительным вершинам, которое также задает множество остальных ребер сети $E_W := (w,m) : w \in W_m, m = 1,\ldots,p$.

Заметим, что $|m' : (m,m') \in E| = r_m$ для всех $m \in M$. Пусть каким-то образом зафиксирована *топология связей* E_M, количество дополнительных вершин, имеющих инцидентные висячие вершины, и количество инцидентных висячих вершин для каждой дополнительной вершины (пункты 1-3). Для краткости будем обозначать через H_M *топологию внутренней сети*, то есть кортеж $< q, (r_1,\ldots,r_q), E_M, p >$. Рассмотрим задачу выбора оптимального разбиения $W_1, \ldots, W_p$ основных вершин по дополнительным вершинам (пункт 4) при фиксированной топологии внутренней сети H_M и фиксированном векторе $k = (k_1, \ldots, k_p)$.

Согласно формуле (3), в древовидной сети для вычисления объема потока через вершину m необходимо из суммарного объема всевозможных потоков вычесть потоки между группами $s_i(m)$ основных вершин, достижимыми через i-е ребро, инцидентное вершине m. При этом достаточно учитывать только ребра, связывающие m с другими дополнительными вершинами (без ограничения общности считаем, что это первые $r_m - k_m$ ребер, перечисляемые в формуле (3)), так как остальные группы состоят из одной основной вершины, и поток через них равен нулю.

Для каждой вершины $m \in M$ введем $n \times (r_m - k_m)$ матрицу $Y_m = (y_{wi}^m)$, столбцы $y_1^m, ..., y_{r_m - k_m}^m$ которой являются индикаторными функциями множеств $s_i(m)$, $i = r_m - k_m$. Тогда формулу (3) для внутренней вершины m в сети H можно записать как

$$x_H(m) = 1_n^T R \cdot 1_n - \sum_{i=1,...,r_m-k_m} y_i^T R y_i = 1_n^T R \cdot 1_n - trY_m^T R Y_m.$$

Введем $n \times p$ матрицу $Y = (y_{wm})$, описывающую распределение основных вершин по дополнительным вершинам; $y_{wm} = 1$ если $w \in W_m$, в противном случае $y_{wm} = 0$. Матрица Y принадлежит множеству Ξ. Выразим матрицы Y_m, $m \in M$ через матрицу Y. Для этого введем $p \times (r_m - k_m)$ матрицы $P_m = (p_{m'i}^m)_{m' \in M, i=1,...r_m-k_m}$, $m = 1, \dots, q$, элемент $p_{m'i}^m$ которых равен единице, если дополнительная вершина $m' \in 1,...,p$ достижима из вершины $m \in M$ через i-ое ребро, инцидентное вершине m. Тогда легко проверить, что $Y_m = YP_m$, $m \in M$. Значит, $x_H(m) = 1_n^T R \cdot 1_n - trP_m^T Y^T RYP_m = 1_n^T R \cdot 1_n - trY^T RYP_m P_m^T$, и задача свелась к поиску

(14) $$\underline{C}_2(H_M, k) := \min_{Y \in \Xi} \sum_{m \in M} c_2(1_n^T R \cdot 1_n - trY^T RYP_m P_m^T).$$

Предположим, что функция $c_2(\cdot)$ выпуклая. Тогда, аналогично тому как это было сделано выше для двухуровневой сети, заменим суммируемые функции в (14) касательными с наклонами $\alpha_1, ..., \alpha_q$:

$$\underline{C}_2(H_M, k) \geq \underline{C}_2^{\cup}(H_M, k) := \sup_{\alpha_1,...\alpha_q \geq 0} \min_{Y \in \Xi} \sum_{m=1}^{q} \{c_2(d_2(\alpha_m)) - \alpha_m d_2(\alpha_m) + \alpha_m(1_n^T R \cdot 1_n - trY^T RYP_m P_m^T)\} =$$

$$= \sup_{\alpha_1,...\alpha_q \geq 0} \{\sum_{m=1}^{q} \left[c_2(d_2(\alpha_m)) - \alpha_m d_2(\alpha_m) + \alpha_m 1_n^T R \cdot 1_n\right] - \max_{Y \in \Xi} trY^T RY \sum_{m=1}^{q} \alpha_m P_m P_m^T\}.$$

Определим $p \times p$ матрицу $P(\alpha, H_M) := \sum_{m=1}^{q} \alpha_m P_m P_m^T$, задаваемую топологией сети H_M и зависящую от вектора параметров линеаризации α. Тогда

$$\max_{Y \in \Xi} trY^T RYP(\alpha, H_M) \leq \max_{Y: Y^T Y = K} trY^T RYP(\alpha, H_M) = \max_{Z: Z^T Z = I_p} trK^{1/2} Z^T RZK^{1/2} P(\alpha, H_M) =$$

$$= \max_{Z: Z^T Z = I_p} trZ^T RZK^{1/2} P(\alpha, H_M) K^{1/2}.$$

Определим матрицу $\tilde{P}(\alpha, H_M, k) := K^{1/2} P(\alpha, H_M) K^{1/2}$. Так как она положительно полуопределена, по следствию 4 имеем $\max_{Y \in \Xi} trY^T RYP(\alpha, H_M) \leq \sum_{i=1}^{p} \lambda_i(R) \lambda_i(\tilde{P}(\alpha, H_M, k))$ и

(15) $$\underline{C}_2^{\cup}(H_M, k) = \sup_{\alpha_1,...\alpha_q \geq 0} \{\sum_{m=1}^{q} \left[c_2(d_2(\alpha_m)) - \alpha_m d_2(\alpha_m) + \alpha_m 1_n^T R \cdot 1_n\right] - \sum_{i=1}^{p} \lambda_i(R) \lambda_i(\tilde{P}(\alpha, H_M, k))\}.$$

Все вышесказанное доказывает следующее утверждение.

Утверждение 7. *Пусть древовидная сеть H состоит из $q \geq 1$ дополнительных вершин, первые p из которых имеют по k_m инцидентных висячих вершин* ($m = 1, \dots, p$), *и фиксирована топология внутренней сети E_M. Тогда* $C_2(H) \geq \underline{C}_2^{\cup}(H_M, k)$.

При этом матрица, доставляющая этот минимум, вычисляется как $Y^* = U_1 U_2^T K^{1/2}$, где U_1 – $n \times p$ матрица, столбцы которой – нормированные на единицу собственные вектора матрицы R, соответствующие p ее наибольшим собственным числам, упорядоченным по убыванию, а U_2 – $p \times p$ матрица, столбцы которой – нормированные на единицу собственные вектора матрицы $\tilde{P}(\alpha, H_M, k)$, также упорядоченные по убыванию соответствующих им собственных чисел.

6. Заключение

Нижняя оценка (15) представляет собой лишь первый шаг к построению нижней оценки затрат древовидной сети, оптимальной на множестве древовидных сетей с заданным количеством дополнительных вершин q и заданными их степенями $r_1, \ldots, r_q$, что нужно для поиска сети H_1^*, минимизирующей затраты $C_2(\cdot)$ на множестве сетей Ω_1.

В качестве следующих шагов необходимо учиться искать минимум выражения(15) по вектору k размеров компонент разбиения, количеству p дополнительных вершин, имеющих инцидентные основные вершины, а также по топологии внутренней сети E_M.

Введем матрицы $D_m(H_M) = E_p - P_m P_m^T$, $m = 1, \ldots, q$, и $D(\alpha, H_M) = \sum_{m=1}^{q} \alpha_m D_m(H_M)$. Если всем q дополнительным вершинам сети H назначить веса $\alpha_1, \ldots, \alpha_q$, то матрица $D(\alpha, H_M) = (d_{mm'})$ – это $p\times p$-матрица, элемент $d_{mm'}$ которой равен сумме весов вершин на пути между вершинами m и m' в H (в силу древовидности H путь единственен). В матрицу D входят только расстояния между p вершинами, имеющими инцидентные висячие вершины.

Тогда, например, с учетом того, что $Y \cdot 1_p = 1_n$, формулу(15) можно переписать как

$$\underline{C}_2^{\cup}(H_M, k) = \sup_{\alpha_1, \ldots \alpha_q \ldots 0} \{\sum_{m=1}^{q} \left[c_2(d_2(\alpha_m)) - \alpha_m d_2(\alpha_m) \right] + \min_{Y \in \Xi} trY^T RYD(\alpha, H_M)\}$$

и выразить непосредственно через собственные числа матрицы расстояний $D(\alpha, H_M)$.

Следовательно, задача поиска оптимальной древовидной сети для фиксированных q, r, p сводится к максимизации взвешенной (здесь веса определяются собственными числами матрицы потоков R) суммы собственных чисел матрицы $K^{1/2} D(\alpha, H_M) K^{1/2}$ на довольно сложном множестве матриц, соответствующих матрицам расстояний древовидных сетей (см. обзор современного состояния теории в [1970]).

Другим перспективным направлением работы является повышение качества нижних оценок и их использование для разработки точных и приближенных алгоритмов поиска оптимальных связывающих сетей для аддитивных функций затрат.

Работа выполнена при поддержке Российского фонда фундаментальных исследований (13-07-00389а).

Список литературы

1. Воронин А.А., Мишин С.П. Оптимальные иерархические структуры. М.: ИПУ РАН, 2003.
2. Гордеев Э.Н., Тарасцов О.Г. Задача Штейнера. Обзор // Дискретная математика. 1993. Т. 5, № 2. С. 3-28.
3. Губко М.В. Математические модели оптимизации иерархических структур. М.: ЛЕНАНД, 2006.
4. Дементьев В.Т., Ерзин А.И., Ларин Р.М. и др. Задачи оптимизации иерархических структур. Новосибирск: Изд-во Новосиб. ун-та, 1996.
5. Мишин С.П. Оптимальные иерархии управления в экономических системах. М.: ПМСОФТ, 2004.
6. Donath W.E., Hoffman A.J. Lower bounds for the partitioning of graphs // IBM J. Res. Dev. 1973. Vol. 17, No. 5. P. 420-425.
7. Rendl F., Wolkowicz H. A projection technique for partitioning the nodes of a graph // Annals of Operations Research. 1995. Vol. 58, No. 3. P. 155-179.
8. Schaeffer S.E. Survey: Graph clustering // Comput. Sci. Rev. 2007. Vol. 1, No. 1. P. 27-64.
9. Stevanović D., Ilić A. Spectral Properties of Distance Matrix of Graphs // In: Distance in Molecular Graphs – Theory. [ed. by I. Gutman, B. Furtula]. Kragujevac: Univ. Kragujevac, 2012, P. 139-176.
10. Winter P. Steiner problem in networks: A survey // Networks. 1987. No. 17. P. 129-167.
11. Wolkowicz H., Zhao Q. Semidefinite programming relaxations for the graph partitioning problem // Discrete Applied Mathematics. 1999. No. 96-97 P. 461-479.