ОСТОВНЫЕ ДЕРЕВЬЯ С МАКСИМАЛЬНОЙ СУММОЙ ПРОИЗВЕДЕНИЙ СТЕПЕНЕЙ ВЕРШИН

К. В. Кухаренко

Белорусский государственный университет, г. Минск;
orlovich@bsu.by
науч. рук. – Ю. Л. Орлович, канд. физ.-мат. наук, доц.

Задача, рассматриваемая в настоящей работе, возникла в области биоинформатики. Исходная постановка задачи такова: имеется множество людей, недавно заболевших некоторым инфекционным заболеванием. Требуется восстановить историю распространения эпидемии, т.е. установить, с кого началась инфекция, и кто кого заразил. Теоретико-графовой моделью рассматриваемой задачи служит остовное дерево графа, описывающего процесс распространения эпидемии, которое обладает максимальной суммой произведений степеней вершин, где суммирование производится по всем рёбрам дерева. В работе рассматриваются структурные и алгоритмические свойства оптимальных остовных деревьев, а также устанавливается вычислительная сложность нахождения соответствующих теоретико-графовых параметров.

Ключевые слова: теоретико-графовая модель; остовное дерево; *s*-метрика; NP-полная задача.

ВВЕДЕНИЕ

Математически рассматриваемую задачу можно сформулировать в терминах теории графов. Дан граф G, вершины которого ассоциируются с людьми, две вершины u и v смежны в графе G тогда и только тогда, когда возможна передача вируса от u к v или наоборот. Требуется найти остовный подграф графа G, который описывает историю распространения эпидемии. Как правило полагается, что один и тот же человек не может быть заражён дважды, поэтому можно считать, что искомый остовный подграф связен и не содержит циклов, т.е. является его остовным деревом.

Немаловажную роль в построении математической модели играет тот факт, что реальные графы, ассоциированные с вирусными заражениями, являются так называемыми «scale-free» графами. Хорошей теоретико-графовой мерой того, насколько граф «scale-free», является s-метрика, введённая в [1]. Графы с наибольшим значением s-метрики наиболее вероятно являются «scale-free», следовательно, возвращаясь к задаче, наиболее вероятно описывают историю распространению инфекции. Остовное дерево графа G с наибольшим значением s-метрики будем называть оптимальным остовом.

Интерес к данной теме был инспирирован исследованиями вирусных заболеваний и процессов распространения эпидемий, проводимыми в CDC (США). В частности, в работе [2] была представлена изложенная выше математическая модель процесса распространения инфекции, и так же было предложено использовать s-метрику.

В настоящей работе установлены оценки для s-метрики и исследовано множество её значений в классе деревьев. Предложены эффективные алгоритмы для нахождения остовных деревьев с максимальной суммой произведений степеней вершин для некоторых классов графов. Доказана NP-полнота задачи распознавания «$\tau_G \geq y$?».

1. ИСПОЛЬЗУЕМЫЕ ПОНЯТИЯ

Теоретико-графовые термины и обозначения, не указанные в данной работе явно, следуют [3]. Рассматриваются конечные неориентированные графы без петель и кратных рёбер. Множество всех остовных деревьев графа G обозначим $SPT(G)$. Связный граф G, содержащий в точности один цикл, называется унициклическим. Вершина u в графе G порядка n называется доминирующей, если $\deg_G(u) = n-1$, т.е. она смежна со всеми остальными вершинами графа. Доминирующим множеством в графе называется множество вершин, такое, что каждая не входящая в него вершина графа смежна с некоторой вершиной этого множества. Связным доминирующим множеством называется множество вершин, которое является доминирующим и порождает связный подграф. Числом связного доминирования $\gamma_c(G)$ называется наименьшая из мощностей его связных доминирующих множеств.

Термины и обозначения теории алгоритмической сложности следуют монографии [4].

2. ПОСТАНОВКА ОПТИМИЗАЦИОННОЙ ЗАДАЧИ

Следуя работе [1], введём для произвольного графа $G = (V_G, E_G)$ s-метрику:

$$s(G) = \sum_{vw \in E_G} \deg_G(v) \deg_G(w).$$

Рассмотрим её как целевую функцию на множестве $SPT(G)$ всех остовных деревьев графа G. Как уже было отмечено выше, остовные деревья с наибольшим значением s-метрики наиболее вероятно описывают историю распространению инфекции. Поэтому оптимизационная задача имеет следующий вид:

$$s(T)=\sum_{vw\in E_T}\deg_T(v)\deg_T(w)\to\max, \qquad (1)$$
$$T\in SPT(G).$$

Определение. Через τ_G обозначим максимум целевой функции

$$\tau_G=\max_{T\in SPT(G)} s(T).$$

Остов $T^*\in SPT(G)$ будем называть оптимальным, если на нём достигается максимум целевой функции, т.е. $\tau_G=s(T^*)$.

В силу специфики рассматриваемой задачи исследование *s*-метрики в данной работе проводится исключительно для класса деревьев, для её обозначения в дальнейшем будем также употреблять параметр *s*(*T*).

3. ОСНОВНЫЕ РЕЗУЛЬТАТЫ

Рассмотрим произвольный граф $G=(V_G,E_G)$. Для оценки параметра τ_G необходимо установить, какие значения принимает *s*-метрика на множестве его остовных деревьев. Верны следующие оценки параметра *s*(*T*).

Утверждение 1. Для любого дерева *T* порядка $n\geq 3$ верна оценка $s(T)\geq 4n-8,$ причём равенство достигается тогда и только тогда, когда $T=P_n$.

Следствие 1. Для любого дерева *T* порядка $n\geq 3$ справедлива оценка $s(T)\geq 2(n+\gamma_c(T))-4,$ где $\gamma_c(T)$ – число связного доминирования дерева *T*.

Следствие 2. Для любого дерева *T* порядка n справедлива оценка $s(T)\geq 2(n+\mathrm{diam}(T))-6,$ где $\mathrm{diam}(T)$ – диаметр дерева *T*.

Утверждение 2. Для любого дерева *T* порядка n верна оценка $s(T)\leq (n-1)^2,$ причём равенство достигается тогда и только тогда, когда $T=K_n$.

Сформулируем теорему для параметра τ_G.

Теорема 1. Для произвольного графа *G* порядка *n* справедливы оценки $4n-8\leq\tau_G\leq(n-1)^2$.

Исследуем множество значение параметра *s*(*T*). Верна следующая теорема.

Теорема 2. Множество значений *s*-метрики для деревьев порядка $n\geq 5$ не покрывает множество целочисленных точек на отрезке $[4n-8,(n-1)^2]$.

Нахождение параметра τ_G для произвольного графа представляется сложной задачей. Тем не менее, результаты, полученные в работе, позволяют эффективно находить параметр τ_G для некоторых классов графов.

Следствие 3. Для графов P_n и C_n достигается нижняя оценка параметра τ_G, т.е. $\tau_G = 4n - 8$.

Следствие 4. Для графа G порядка n верхняя оценка параметра τ_G достигается тогда и только тогда, когда G содержит доминирующую вершину.

Следствие 5. Для произвольного графа G порядка n проверка достижимости верхней границы параметра τ_G эквивалента задаче определения, содержит ли граф доминирующую вершину, и, следовательно, выполняется за время $O(n)$.

Теорема 3. Для унициклического графа G параметр τ_G и оптимальный остов T^* можно найти за время $O(n)$.

Для изучения вычислительной сложности задачи (1) важная следующая лемма о структуре s-метрики.

Лемма 1. Для любого дерева T порядка n верно равенство $s(T) = P_1(T) + 2 \cdot P_2(T) + P_3(T)$, где $P_3(T)$ – число всех простых цепей длины три между вершинами дерева T, $P_2(T)$ – число всех простых цепей длины два и $P_1(T) = n - 1$ – число рёбер в T.

Сформулируем для задачи (1) соответствующую ей задачу распознавания: Верно ли, что для графа G и положительного целого числа y выполняется соотношение $\tau_G \geq y$? Имеет место следующая теорема.

Теорема 4. Задача распознавания «$\tau_G \geq y$?» NP-полна.

Схема доказательства теоремы 4 позволяет установить аналогичное утверждение и в специальном случае, когда граф G двудольный.

Следствие 6. Задача распознавания «$\tau_G \geq y$?» является NP-полной в классе двудольных графов.

Библиографические ссылки

1. *Li, L.* Towards a theory of scale-free graphs: definitions, properties and implications / L. Li, D. Andersen, J.C. Doyle, W. Willinger // Internet Mathematics. 2005. Vol. 2, № 4. P. 431–523.
2. *Skums, P.* QUENTIN: reconstruction of disease transmissions from viral quasispecies genomic data / P. Skums, A. Zelikovsky, R. Singh, C. Sahinalp // Bioinformatics. 2018. Vol. 34. P. 163–170.
3. Лекции по теории графов / В.А. Емеличев [и др.]. М.: Наука, 1990. 384 с.

4. *Garey, M.R.* Computers and Intractability: A Guide to the Theoty of NP-Completeness / M.R. Garey, D.S. Johnson. New York, NY, USA: W.H.Freeman & Co., 1979.