МОДИФИКАЦИЯ АЛГОРИТМА SPREAD ДЛЯ РАСПРОСТРАНЕНИЯ СООБЩЕНИЯ В СЛУЧАЙНОМ ГРАФЕ

Рыжов М. С.[1]

(*ФГБУН Институт проблем управления им. В.А. Трапезникова РАН, Москва*)

Работа посвящена распространению сообщения в ненаправленном случайном графе. Предполагается, что одно сообщение, содержащееся вначале в единственном начальном узле, на каждом шаге модифицированного алгоритма SPREAD отправляется другим узлам. Цель работы состоит в исследовании необходимого количества шагов алгоритма, чтобы сообщение было доставлено заданному заранее количеству узлов. В работе приводится оценка сверху этого количества шагов.

Ключевые слова: распространение сообщений, случайный граф, проводимость, изопериметр.

1. Введение

Задача распространения сообщений имеет важное применение в задачах распределённых вычислений [5], а также в исследовании распространении слухов и информации в социальных системах [3]. В работе [4] было проведено исследование экстремального индекса сообщества узлов и минимального времени распространения сообщения до одного узла до всех остальных узлов ненаправленного графа.

Одним из основных изучаемых алгоритмов в задачах распространения является алгоритм SPREAD [1, 5]. На каждом шаге алгоритма выбирается узел i ненаправленного графа, равномерно среди всех узлов. Далее с заданной вероятностью $P_{i,j}$ среди связанных с ним узлов выбирается узел j, который получит все сообщения от узла i. В [5] приводится доказательство верхней границы числа шагов алгоритма SPREAD, которое необходимо для обмена каждого узла ненаправленного графа своим сообщением с остальным графом.

---

[1] *Максим Сергеевич Рыжов, м.н.с.* (*maksim.ryzhov@frtk.ru*).

В текущей работе исследуется модификация алгоритма SPREAD для передачи одного сообщения на заданное количество узлов ненаправленного графа. Узел i для передачи сообщения другим узлам в предложенной модификации выбирается только среди тех узлов, которые на текущей момент обладают сообщением. В работе приводится доказательство верхней границы числа шагов модифицированного алгоритма, необходимое для доставки одного сообщения заданному количеству узлов.

Статья организована следующим образом. В разделе 2 представлен обзор и постановка задачи распространения информации. В разделе 3 представлен главный результат о времени распространения одного сообщения, который потом продемонстрирован на сгенерированном графе (раздел 5). Также в разделе 4 приводится подход к оцениванию проводимости или изопериметра графа. В конце работы приводится итоговое заключение о полученном результате (раздел 6).

2. Задача распространения сообщений

Опишем идею алгоритма распространения, предложенного в [5] для неориентированного графа $G = (V, E)$. Здесь V и E – это наборы узлов и ребер графа соответственно. Один из узлов может инициировать передачу сообщений по тику глобальных часов, которые моделируются как процесс Пуассона с параметром $n = |V|$ [1, 5]. Пусть $k \geq 0$ обозначает номер шага алгоритма или номер тика, по которому не более чем один узел может получить сообщения, связываясь с другим узлом. Изначально $k = 0$.

В случае полной передачи сообщений вводится необходимое количество тиков

$$K(\delta) = inf\{k \geq 0 : \sum_{i=1}^{n} P(\{S_i(k) \neq V\}) \leq \delta\},$$

при котором каждый узел смог доставить своё сообщение до всех остальных узлов с вероятностью $1 - \delta \in (0, 1)$. $S_i(k)$ обозначает набор узлов, которые получили сообщение m_i от узла i на момент тика часов k. Передача сообщений осуществлялась с помощью алгоритма SPREAD [1, 5]. На каждый тик k равномерно среди всех узлов выбирается узел i. Независимо от прошлого выбора узел i выбирает связанный узел j с вероятностью $P_{i,j} = 1/D_i$, и узел i отправляет все сообщения, которые у него есть, узлу j, так что $M_j(k + 1) = M_i(k) \cup M_j(k)$, где $M_i(k)$ – набор полученных узлом i сообщений при тике k. Здесь D_i – степень узла графа, количество его связей. Для SPREAD алгоритма доказывается следующая лемма (лемма 4, [5]):

Лемма 1. Для любого $\delta \in (0, 1)$ определим $K(\delta)$. Тогда $K(\delta) = O\left(n\dfrac{log(n) + log(\delta^{-1})}{\Phi(G)}\right)$.

Здесь $\Phi(G)$, $0 \le \Phi(G) \le 1$, – это проводимость графа, или изопериметр, описывающий насколько связан ненаправленный граф по меньшей мере [1, 5, 6]:

$$(1)\quad \Phi(G) = \min_{S\subset V, 0<|S|<n/2} \frac{\sum_{i\in S, j\notin S} P_{i,j}}{|S|} = \min_{S\subset V} \frac{\sum_{i\in S, j\notin S} P_{i,j}}{min(|S|, |V| - |S|)}.$$

При $\Phi(G) = 0$ граф не является связанным, т.е. существуют узлы, не имеющие связей с остальным графом.

3. Модификация алгоритма SPREAD с известными получателями сообщений

Рассмотрим модификацию алгоритма SPREAD, в которой узел, инициирующий передачу сообщений дальше, выбирается не среди всех узлов графа, а только среди тех, что уже обладают сообщением. Это позволит избежать ситуации, когда выбранный узел не обладает информацией для передачи сообщений. Поэтому модифицированный алгоритм потребует меньшего числа шагов для доставки сообщений, потому что будет меньше ситуаций, когда количество уже получивших узлов осталось неизменным.

Пусть в дальнейшем будет рассматриваться передача одного сообщения из одного узла до заданного количества узлов ненаправленного графа $G = (V, E)$, $N = |V|$. Определим необходимое количество тиков часов $K(n, \delta) = inf\{k\ge 0$: $P(S(k) < n \le \delta\}$, при котором сообщение не было получено $n \le N$

узлами с вероятностью $1 - \delta \in (0, 1)$. Здесь обозначим $S(k)$ как группу узлов, которая имеет сообщение в момент тика часов k, $S(0) = 1$. Докажем следующую лемму:

Лемма 2. Для любого $\delta \in (0, 1)$, $n > 1$, определим количество шагов алгоритма $K(n, \delta)$. Для модифицированного SPREAD алгоритма и ненаправленного связанного графа $G = (V, E)$, $N=|V|$,

$$K(n,\delta) = \begin{cases} O\left(\dfrac{n-\delta}{\delta\Phi(G)}\right), n < N/2, \\ O\left(\dfrac{(n-N/2)(n-1)}{\delta\Phi(G)(N+1-n)} + \dfrac{N/2-\delta}{\delta\Phi(G)}\right), N/2 \le n \le N; \end{cases} \quad (2)$$

где $\Phi(G)$ – проводимость или изопериметр (1) ненаправленного графа G.

Доказательство. Чтобы доказать лемму, рассмотрим изменения набора узлов $S(k)$ с течением временем. Как в доказательстве леммы 1, рассмотрим возможные ситуации во время передачи сообщения на $k + 1$ шаге алгоритма. Возможны следующие случаи: $|S(k + 1)| - |S(k)| = 0$ или $|S(k + 1)| - |S(k)| = 1$. Последний случай возможен только если алгоритм выбрал узел j для получения сообщения и тот его не содержит. Учитывая это и то, что передающий сообщение узел i, выбирается среди узлов с ним, получаем

$$E[|S(k+1)| - |S(k)| \mid S(k)] = \sum_{i \in S(k), j \notin S(k)} \frac{1}{|S(k)|} P_{i,j}.$$

Далее рассмотрим несколько случаев. Пусть $1 \le S(k) \le N/2$ для всех возможных k, тогда

$$E[|S(k+1)| - |S(k)| \mid S(k)] = \sum_{i \in S(k), j \notin S(k)} \frac{1}{|S(k)|} P_{i,j} \ge \Phi(G),$$

где мы воспользовались определением проводимости (1). В результате получаем

$$E[|S(k+1)| \mid S(k)] \ge |S(k)| + \Phi(G).$$

Далее доказательство опирается на доказательство леммы 1. Пусть $Z(k) = S(k) - \Phi(G)k - 1$. Определим такой момент

$L_1 = \inf\{k: | S(k)| > n\}$ и $L_1 \wedge k = \min(L_1, k)$. Таким образом неравенство, полученное для условного матожидания, определяет $Z(L_1 \wedge k)$ как субмартингал. Чтобы проверить это, докажем основное свойство субмартингала, что $E[Z(L_1 \wedge (k + 1)|S(k)] \geq$ $\geq Z(L_1 \wedge k)$. Докажем для случая, когда $|S(k)| \leq n$:

$$E[Z(L \wedge (k+1)) \mid S(L_1 \wedge k)] = E[| S(L_1 \wedge (k+1)) | \mid S(L_1 \wedge k)] -$$
$$-E[\Phi(G)(L_1 \wedge (k+1)) + 1 \mid S(L_1 \wedge k)] \geq$$
$$\geq \Phi(G) + E[| S(L_1 \wedge k) | \mid S(L_1 \wedge k)] -$$
$$-E[\Phi(G)((L_1 \wedge k) + 1) + 1 \mid S(L_1 \wedge k)] =$$
$$= E[| S(L_1 \wedge k) | \mid S(L_1 \wedge k)] -$$
$$-E[\Phi(G)(L_1 \wedge k) + 1 \mid S(L_1 \wedge k)] = Z(L_1 \wedge k).$$

Поскольку $Z(L_1 \wedge k)$, для любого $k > 0$ справедливо $E[Z(L_1 \wedge 0] \leq E[Z(L_1 \wedge k)]$ и $E[1 + \Phi(G)(L_1 \wedge k)] \leq E[|S(L_1 \wedge k)|]$, при условии $Z(L_1 \wedge 0) = Z(0) = 0$. Используя это факт, получаем

$$E[1 + \Phi(G)(L_1 \wedge k)] \leq n,$$

где мы воспользовались фактом, что сообщение должно быть распространено на n узлов. При $k \to \infty$

$$E[1 + \Phi(G)L_1] \leq n.$$

Используя неравенство Маркова, выпишем неравенства $P(L_1 > k_1) = P(1 + \Phi(G)L_1 > 1 + \Phi(G)k_1) < \dfrac{n}{1 + \Phi(G)k_1} < \delta,$ где предполагается $P(L_1 > k_1) = P(S(k) < n)$. При $k_1 = \dfrac{n - \delta}{\delta\Phi(G)}$ будет выполнено условие $P(|S(k)| \neq n) < \delta$, а значит, ограничение для $K(n, \delta)$ представимо в виде

$$K(n, \delta) = O\left(\frac{n - \delta}{\delta\Phi(G)}\right).$$

Если $N > |S(k)| \geq N/2$, начиная с некоторого k, то

$$E[| S(k+1) | - | S(k) | \mid S(k)] = \frac{N - | S(k) |}{| S(k) |} \frac{\sum\limits_{i \in S(k), j \notin S(k)} P_{i,j}}{N - | S(k) |} \geq$$
$$\geq \frac{N - | S(k) |}{| S(k) |} \Phi(G) \geq \frac{N - n + 1}{n - 1} \Phi(G).$$

Введя $L_2 = \inf\{k\colon |S(k)| > n\}$ и пользуясь свойствами субмартингала, аналогично случаю $1 \le S(k) \le N/2$, можно получить

$$E\left[\frac{N-n+1}{n-1}\Phi(G)(L_2 - k_{N/2})\right] \le n - N/2,$$

где $k_{N/2} = K(N/2, \delta)$. Снова воспользуемся неравенством Маркова:

$$P(L_2 > k_2) = P\left(\frac{N-n+1}{n-1}\Phi(G)(L_2 - k_{N/2}) > \frac{N-n+1}{n-1}\Phi(G)(k_2 - k_{N/2})\right) <$$

$$< \frac{n - N/2}{\frac{N-n+1}{n-1}\Phi(G)(k_2 - k_{N/2})} < \delta.$$

Неравенство будет выполнено при $k_2 = \dfrac{(n-1)(n-N/2)}{\delta(N-n+1)\Phi(G)} + k_{N/2}$,

что значит

$$K(n,\delta) = O\left(\frac{(n-N/2)(n-1)}{\delta\Phi(G)(N+1-n)} + \frac{N/2-\delta}{\delta\Phi(G)}\right).$$

4. Оценивание проводимости

В доказанной выше лемме 2 фигурирует значение проводимости графа $\Phi(G)$. Поиск точного её значения является NP-сложной задачей, если не рассматривается конкретный тип ненаправленного графа. Поэтому требуется оценивать примерное значение.

Для проводимости $\Phi(G)$, которую также называют константой Чигера h_c, справедлива следующая теорема.

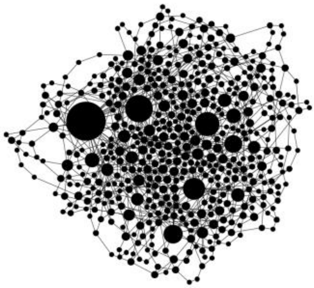
Теорема 1. *Пусть* $0 = \lambda_0 \le \lambda_1 \le \ldots \le \lambda_{n-1}$ *– собственные значения нормализованной матрицы Лапласа* $L = D^{-1}A$ *для ненаправленного графа* G*, где матрица* A *– матрица связей графа, а матрица* D *– матрица степеней узлов. Для константы Чигера* h_c *справедливо* $\sqrt{1-h_C^2} \le \lambda_1 \le 2h_C$.

Поиск собственных значений является простой задачей, а поэтому значение проводимости можно оценить по наименьшему положительному собственному значению Лапласа:

$$(3)\quad \hat{\Phi}(G) = \frac{\sqrt{1-\lambda_1^2} + 2\lambda_1}{2}.$$

5. Моделирование

Для проверки был сгенерирован граф с помощью модели Thorny Branching Tree (TBT) с заданным распределением числа связей (рис. 1 справа). Граф содержит 500 узлов и его проводимость, которая оценена по (3), равна $\Phi(G)$=0,4735. Для графа было проведено моделирование модифицированного алгоритма SPREAD (раздел 3). Полученные значения числа шагов алгоритма с оцененной верхней границей (2) представлены на рис. 1, слева. Из рисунка видно, что верхняя граница примерно описывает поведение алгоритма при различных значениях n.



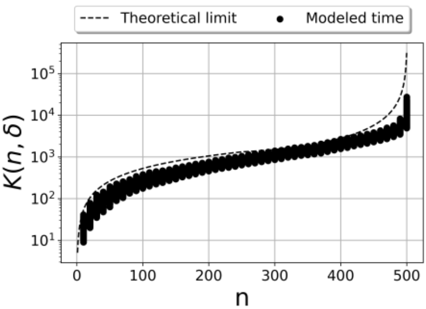


Рис. 1. Смоделированный TBT граф (*слева*), N = 500*, размеры узлов пропорциональны их степеням* D_i*. График количества шагов модифицированного алгоритма SPREAD, необходимого для доставки сообщения* n *узлам, против числа* n (*справа*)*. Точками обозначены значения полученные моделированием, прерывистой линией обозначена верхняя границы* $K(n, \delta)$ (2), δ*=0,05*

Из рисунка и из формулы (2) можно сделать вывод, что алгоритм для малого числа узлов n из общего числа N работает быстрее, чем при больших n. Это обусловлено, тем что на первых шагах алгоритма на каждом шаге алгоритма у любого узла из $S(k)$ с большой вероятностью возможно найти соседний узел без сообщения. Верно и обратное, что при большом количестве узлов, которые должны получить сообщения, алгоритм требует всё больше числа шагов. Это происходит, потому что большинство узлов из $S(k)$ уже не имеют связанных с ними узлов, еще не получивших сообщение. Это приводит к росту числа шагов алгоритма, которые происходят без увеличения множества узлов $S(k)$, уже получивших сообщений.

6. Заключение

В работе предложен модифицированный алгоритм SPREAD, в котором общение инициируется узлом, который выбирается среди уже получивших сообщение. Для модификации получена оценка верхней границы числа шагов, необходимых для доставки одного сообщения от одного узла до заданного количества других. В отличие от известного алгоритма SPREAD, граница для модификации имеет линейный и квадратичный участки зависимостей от числа узлов.

Литература

1. CENSOR-HILLEL K., SHACHNAI H. *Partial Information Spreading with Application to Distributed Maximum Coverage* // Proc. of the 29th ACM Symposium on Principles of Distributed Computing (PODC'10), ACM, New York, USA. – 2010. – P. 161–170.
2. CHUNG F. *Spectral Graph Theory* // CBMS Regional Conference Series in Mathematics. – 1997. – Vol. 92. – American Mathematical Society, Providence, RI, 1997.
3. LIND P., SILVA L., ANDRADE JR., HERRMANN H. *Spreading gossip in social networks* // Physical review. E, Statistical, Nonlinear, and Soft Matter Physics. – 2007.
4. MARKOVICH N.M., RYZHOV M.S. *Leader Nodes in Communities for Information Spreading* // LNCS. – 2020. – Vol. 12563. – P. 475–484.

5. MOSK-AOYAMA D., SHAH D. *Computing separable functions via gossip* // Proc. of the 25th ACM Symposium on Principles of Distributed Computing (PODC'06), ACM, New York, USA. – 2006. – P. 113–122.
6. NEWMAN M.E.J. *Networks: An Introduction.* – Oxford University Press, Second edition, 2018.
7. VOLKOVICH Y., LITVAK N., DONATO D. *Determining Factors Behind the PageRank Log-Log Plot, Algorithms and Models for the Web-Graph*. – Springer, Berlin, Heidelberg, 2007. – P. 108–123.

MODIFIED SPREAD ALGORITHM FOR MESSAGE PROPAGATION IN A RANDOM GRAPH

Maksim Ryzhov, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow (maksim.ryzhov@frtk.ru).

Abstract: The paper is devoted to the propagation of a message in an undirected random graph. It is assumed that one message, contained initially in a single initial node, is sent to other nodes at each step of the modified SPREAD algorithm. The main purpose is to study the necessary number of steps of the algorithm so that the message is delivered to a predetermined number of nodes. The paper provides an upper bound for this number of steps.

Keywords: information spreading, random graph, conductance, isoperimetric number.

УДК 519.2
ББК 22.172