

Н. М. МАРКОВИЧ, М. С. РЫЖОВ
Институт проблем управления им. в.А.Трапезникова РАН, Москва

РАСПРОСТРАНЕНИЕ ИНФОРМАЦИИ В ЭВОЛЮЦИОНИРУЮЩИХ НЕОДНОРОДНЫХ СЕТЯХ

Рассматривается задача распространения одного сообщения от одного узла направленной неоднородной сети фиксированному числу узлов этой сети. Для этой цели схемы линейного предпочтительного присоединения узлов, используемые ранее для эволюции сети, используются как метод распространения сообщения и сравниваются с известным алгоритмом SPREAD по скорости распространения сообщения. Показано, что сообщества с наименьшими хвостовыми индексами числа выходящих связей и Пейджранга распространяют информацию быстрее других. Результаты получены эмпирически на модельных и реальных графах.

Введение. Предлагаемый доклад посвящен важной практической задаче распространения информации в случайных неоднородных направленных сетях. Распространение сообщений рассматривалось в ненаправленных сетях в работах [1], [2]. Среди применений можно назвать параллельные вычисления в компьютерных сетях, распространение сообщений в социальных сетях и заражений в популяциях. Выбор лидирующего сообщества сети и сравнение распространения сообщений этим сообществом и другими сообществами рассматривался в [3] на примере однородных геометрических графов. Суть метода SPREAD, предложенного в [1] для полного обмена сообщениями между всеми узлами сети, и в [2] для передачи сообщения части узлов, состоит в следующем. Предполагая, что узлы сети имеют асинхронные часы, сообщение между узлами может передаваться только по сигналу общих часов. Эти тики подаются согласно пуассоновскому процессу. Тогда время, необходимое для передачи сообщения, может быть вычислено, как сумма случайного числа экспоненциально распределенных случайных величин.

С другой стороны, на практике часто наблюдается рост сетей вследствие эволюции, т. е. присоединения новых узлов и их связей к существующим узлам. Эволюция сетей моделируется с помощью предпочтительного присоединения (ПА) нового узла к существующему узлу пропорционально числу его связей, либо путем кластерного присоединения [4]. Эволюция начинается с начального графа, содержащего не менее одного узла. Для направленных сетей (и описывающих их графов) в работе [5] предложены три схемы линейного ПА α, β, γ , где вновь создаваемые ребра случайного графа направлены от существующего узла к новому и наоборот, либо эти ребра добавляются между существующими узлами, если новый узел не присоединяется на следующем шаге эволюции.

Доклад содержит результаты работы [6]. Предложено несколько новых подходов.

Во-первых, схемы линейного ПА β и γ предложено использовать для продвижения сообщения от одного узла к фиксированному числу узлов сети. При этом информация может быть передана от узла i , обладающего сообщением, к узлу j только, если между узлами создается направленное ребро от i к j . Эти схемы ПА сравниваются с алгоритмом SPREAD по минимальному числу шагов эволюции, необходимых, чтобы распространить одно сообщение от узла, им обладающего, среди фиксированного числа узлов сети. Изучено влияние выбора начальной сети (подграфа), с которой начинается эволюция, на скорость распространения.

Во-вторых, алгоритм SPREAD, изначально предложенный для ненаправленных графов, адаптирован к направленным. Все результаты получены для неоднородных сетей, где распределения входящих и выходящих связей узлов могут быть различными.

Новые подходы исследуются эмпирически для однородных модельных графов и неоднородных реальных графов из базы данных Берклей–Стенфорд [7] и временных графов [8]. Показано, что сообщества узлов с наименьшими хвостовыми индексами числа выходящих связей и Пейджрангов могут распространять сообщение быстрее. Полученные результаты согласуются с практикой. Показано, что схемы ПА являются лучшими распространителями, чем алгоритм SPREAD, если пропорция вновь созданных связей, направленных от новых узлов, не обладающих сообщением, к существующим узлам мала. Важным результатом является обширное ис-

следование по классификации реальных временных графов (т. е. графов, у которых ребра образуются в определенные моменты времени) по числу шагов эволюции, необходимых для распространения сообщения среди фиксированного числа узлов сети.

Основные результаты

Сравним способность распространять информацию для ПА и алгоритма SPREAD. Пусть число шагов эволюции k ограничено: $k \leq K'$. Определим число тиков часов, требуемых для доставки сообщения от какого-то узла к n узлам с вероятностью, не меньше, чем $1 - \delta$, как

$$K^*(n, \delta) = \inf\{0 < k \leq K' : P(|S(k) = n| > 1 - \delta) > 1 - \delta\}, \delta \in (0, 1),$$

где $S(k)$ – число узлов, получивших сообщение на шаге k . Результаты сравнения алгоритмов ПА и SPREAD показаны на рис. 1 и 2. Параметры ПА α и γ показывают пропорции вновь создаваемых связей, направленных от нового узла к существующему и наоборот, соответственно, а β – пропорцию новых связей между уже существующими узлами. Из рис. 1 следует, что ПА быстрее распространяет сообщение, если $\gamma > 0.51$ или $\beta > 0.6$. Случай $\gamma + \beta > 1$ не рассматривался, так как $\alpha + \beta + \gamma = 1$.

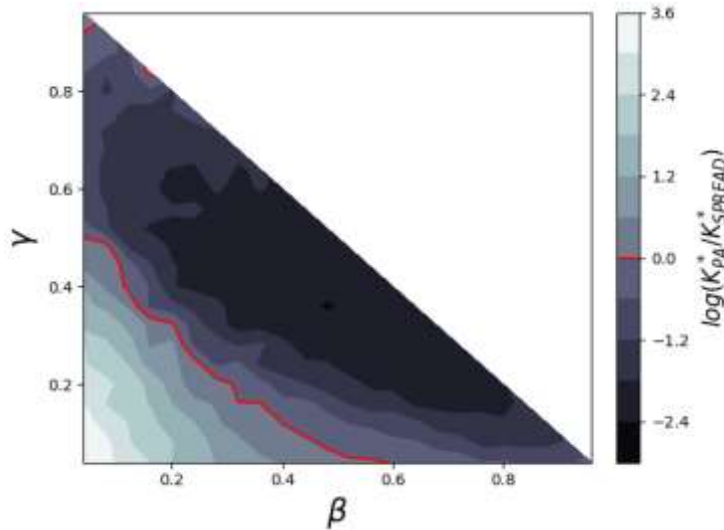


Рис. 1. Зависимость $\log(K_{ПА}^* / K_{SPREAD}^*)$ с минимальным числом шагов $K_{ПА}^*$ и K_{SPREAD}^* , необходимых для доставки одного сообщения к 100 узлам, от параметров ПА β и γ . Усреднение проведено по 100 модельным графам. Красная линия указывает случаи $K_{ПА}^* = K_{SPREAD}^*$, слева от линии $K_{ПА}^* > K_{SPREAD}^*$, справа – $K_{ПА}^* < K_{SPREAD}^*$

Влияние выбора начальной сети, с которой начинается эволюция, исследовалось на реальном графе, представляющем вебстраницы из доменов berkely.edu, stanford.edu [7]. Граф был разбит на сообщества, которые экзаменовались в качестве начальной сети. Выяснилось, что сообщества с наиболее тяжелыми хвостами Пейджрангов и числа выходящих связей быстрее распространяют сообщения, т. е. для них величина K^* наименьшая.

Эффективность ПА и SPREAD сравнивалась для реальных временных графов, рис. 2. Среди них графы сообщений и комментариев с вебсайтов (sx-mathoverflow, sx-askubuntu, CollegeMsg), графы транзакций биткоинов (soc-sign-bitcoin-otc, soc-sign-bitcoin-alpha) и графы коммуникаций по e-майлу (email-Eu, email-Eu-Dept1 and others). Все параметры ПА оценивались методом Snapshot [5] ($\hat{\alpha}^{SN}, \hat{\beta}^{SN}, \hat{\gamma}^{SN}, \hat{\Delta}_{in}^{SN}, \hat{\Delta}_{out}^{SN}$). Для всех графов, за исключением графов биткоин, ПА доставляет сообщение быстрее, чем SPREAD, рис. 2 (слева). Для графов биткоин хвостовые индексы выходящих связей меньше единицы, $\hat{\alpha}_{in} < 1$, рис. 2 (справа). Это означает, что число выходящих связей имеет бесконечную дисперсию по свойству полиномиального распределения. Это является следствием доминирования α -схемы ПА, т.е. создания с большей вероятностью новых связей, направленных от новых узлов без сообщения к старым. В этом случае

сообщение не передается, и число узлов, обладающих этим сообщением, не увеличивается на шаге эволюции.

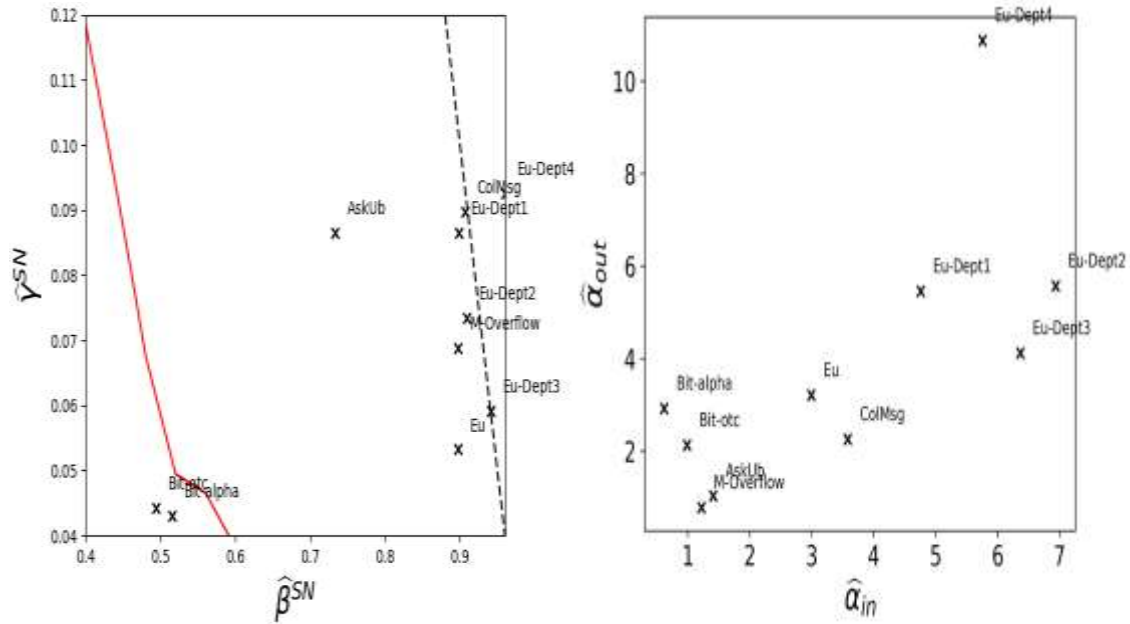


Рис. 2. Минимум $\log(K_{PIA}^* / K_{SPREAD}^*)$ для временных графов, где красная линия указывает случай $K_{PIA}^* = K_{SPREAD}^*$, слева от линии $K_{PIA}^* > K_{SPREAD}^*$, справа – $K_{PIA}^* < K_{SPREAD}^*$, условие выполняется для области слева от пунктирной линии (слева), хвостовые индексы $\hat{\alpha}_{in}$ и $\hat{\alpha}_{out}$ для числа входящих и выходящих связей (справа)

Заключение. Новизна работы заключается в применении схем ПА для распространения информации. Алгоритм SPREAD пересмотрен для направленных графов. Оба метода ПА и SPREAD применены к неоднородным графам, которые обычно возникают на практике. Сообщение от одного узла передается фиксированному числу узлов сети. Распространение сообщения исследуется эмпирически на модельных однородных графах и неоднородных реальных временных графах и графе Беркли–Станфорд. Неоднородность означает, что узлы графов могут иметь различные распределения входящих и выходящих связей. ПА и алгоритм SPREAD сравниваются на направленных графах, возможно содержащих селф-циклы и несколько связей между узлами. Эти графы генерируются с помощью эволюции из небольших начальных графов посредством присоединения узлов и ребер по ПА с различными наборами параметров. К полученным графам применялся алгоритм SPREAD и сравнивался с ПА. На этих модельных однородных графах можно заключить, что ПА может быть лучшим распространителем, чем SPREAD, для наборов параметров ПА, где пропорция новых ребер, направленных от новых узлов к старым, мала.

Рассматривая неоднородные реальные графы, было получено, что сообщества узлов с наименьшими хвостовыми индексами для выходящих связей и Пейджрангов могут распространять информацию быстрее, чем узлы других сообществ. Временные графы классифицированы по числу шагов эволюции, необходимых, чтобы доставить сообщение к конечному числу узлов. Для большинства временных графов ПА оказалось лучшим распространителем, чем SPREAD.

Направления дальнейших исследований связаны с использованием кластерных присоединений и их сравнения с ПА.

Работа проводилась при поддержке гранта РФФ 22-21-00177

ЛИТЕРАТУРА

1. 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.
2. 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.
3. Markovich N.M., Ryzhov M.S. Leader Nodes in Communities for Information Spreading, LNCS 2020. 12563, P. 475–484.
4. Bagrow J.-P., Brockmann D. Natural Emergence of Clusters and Bursts in Network Evolution. *Physical Review X*. 2013. P. 021016.
5. Wan P., Wang T., Davis R. A., Resnick S.I. Fitting the linear preferential attachment model. *Electron. J. Statist.*, 2017. Vol. 11. №2, P. 3738–3780.
6. Markovich N.M., Ryzhov M.S. Information Spreading and Evolution of Non-Homogeneous Networks. *Adv Syst Sci Appl.* (To appear)
7. Leskovec J., Lang K., Dasgupta A., Mahoney, M. Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters. *Internet Mathematics*. 2009. Vol. 6. №1. P. 29–123.
8. Leskovec J., Krevl A. SNAP Datasets: Stanford Large Network Dataset Collection. 2014. <http://snap.stanford.edu/data>.

N.M.Markovich, M.S.Ryzhov, (V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow)

Spreading information in evolved heterogeneous networks

The problem of spreading a message from one node of a directed heterogeneous network to a fixed number of nodes of this network is considered. For this purpose, the schemes of linear preferential attachment of nodes used earlier for the evolution of the network are used as a method of message propagation and are compared with the well-known SPREAD algorithm in terms of the speed of message propagation. It is shown that communities with the lowest tail indices of the number of outgoing links (the out-degree) and the PageRank spread information faster than others. The results are obtained empirically on simulated and real graphs.