

简单、快速的确定型八卦传播算法

赖文星 · 17210240130
复旦大学 · 计算机科学技术学院

2018 年 6 月

1 问题介绍

八卦 (*gossip*, 流言、小道消息) 在人类进化中扮演了重要角色, 同时也能促进社会稳定 [Dunbar, 2004]; 也由有实验表明, 八卦促进了社会合作 [Feinberg et al., 2012]。关于八卦, 我们通常能够观察到这样一种现象, 即当你把秘密告诉你的好友, 不久之后该秘密似乎就众人皆知了——八卦何来如此强大的传播能力? 在 Chierichetti 等人的论文指出, 人群中随机地传播八卦, 只需要将经过 $O(\frac{\log n}{\Phi})$ 轮传播 (这里 Φ 表明图的电导率, *conductance*), 八卦就有很高的概率变得众人皆知 [Chierichetti et al., 2018]。同时, 在未知网络中广播信息, 也是计算机网络中的一个基本问题; 而八卦传播算法 (*gossip algorithms*), 即每个节点每次只能联络一个邻居节点的模型中, 已经有了广泛的研究。例如, 洪泛 (*flooding*) 的方法, 即以此地向邻居节点发送、交换信息被证明是高效的, 并且可以利用该方法使得八卦传播轮数与图的电导率无关 [Censor-Hillel et al., 2012]。

在本文中我们将介绍 Haeupler 提出的第一个确定的与图的电导率无关的八卦传播方法¹[Haeupler, 2015], 并将其与随机的 PULL-PUSH 模型进行对比。

1.1 结果

Haeupler 的文章提供了一种快速、简单、高效且**确定**的解决局部广播问题的八卦传播算法。

定理 1.1. 对于任意的 $k \in \mathbb{N}$, 存在一个确定的八卦传播算法, 在任意的 n -节点图上运行 $2(k \log n + \log^2 n)$ 轮即可解决 k -局部广播问题, 即每个节点与其距离不超过 k 的节点完成八卦交换。

取 $k = D$, 即可立即得到以下推论:

推论 1.1.1. 存在一个确定的八卦传播算法, 其在任意 n -节点且直径为 D 的图上运行 $2(D \log n + \log^2 n)$ 轮即可解决全局广播问题。

¹该文章获得了 SODA13 最佳学生论文, 并刊载于 2015 年第 6 期 *Journal of the ACM*。

Haeupler 的该文最重要的贡献在于, 该作者提出了第一个确定的解决广播问题的算法, 且比 Censor-Hillel 等人的算法解决 $k = 1$ 的局部广播问题快 $\Theta(\log n)$, 对于 $k = \log n$ 则要快 $\Theta(\log^2 n)$; 并且该算法是确定的, 能够给运行时间以确定的保证。另外, 该算法更简单、自然, 对其分析也十分直接, 具有美感。

1.2 八卦传播算法与局部广播问题

在本小节中我们将简单介绍问题的模型以及相关的定义。

基本概念 关于图的基本概念, 请参阅 [Diestel, 2017]。此外, 我们再次特别指出图的电导率的定义。在此之前, 我们定义 $\text{vol}(S) = \sum_{v \in S} \deg(v)$, 即节点集合 S 中节点的度数和。

定义 1.1 (电导率 [Jerrum and Sinclair, 1989]). 在图中非空节点结合 S 的电导率为

$$\Phi(S) = \frac{|\text{cut}(S, V \setminus S)|}{\text{vol}(S)}$$

对于图的电导率, 我们定义它为

$$\Phi = \min_{S \subseteq V, 0 < \text{vol}(S) \leq \frac{\text{vol}(V)}{2}} \frac{|\text{cut}(S, V \setminus S)|}{\text{vol}(S)}$$

网络 我们记无向图 $G = (V, E)$ 的节点数为 $n = |V|$, 边数为 $m = |E|$, 直径为 D 。对于每个节点 $v \in V$, 我们记 $\Gamma(v) = \{u | \{u, v\} \in E\} \cup \{v\}$; 类似地, 我们记 $\Gamma^k(v) = \{u | \text{there is a path of length } k \text{ from } v \text{ to } u\} \cup \{v\}$ 。

八卦传播 这里, 我们的模型中每个节点**同步地**按轮 $t \in \{0, 1, \dots\}$ 传播八卦信息。与 Censor-Hillel 等人定义的模型一致, 我们对于每次信息传播的量不作限制 (因为在现代网络中, 信息量相对较少的信息传播时间并非主要的瓶颈), 并且我们假设每次传播都是**交换**信息 (而非 PULL-PUSH 中的仅仅是单方向的传播)。这里, 我们也假设最开始每个节点不知道除了其邻居节点的之外的任何信息。

局部广播问题 我们现在定义我们所讨论的问题。

定义 1.2 (全局广播问题). 每个节点 v 以八卦 r_v 开始, 其目标在于每个节点都知道所有的八卦。

在该文章中, 作者主要关注于 k -局部广播问题:

定义 1.3. 每个节点 v 以八卦 r_v 开始, 其目标在于每个节点 v 都知道 $\Gamma^k(v)$ 中所有节点的八卦。

我们关注 k -局部广播问题的原因在于:

- 该问题是全局广播问题的推广; 实际上对于任意的 $k > D$, k -局部广播问题等同于全局广播问题。

- 在之前的许多工作中， k -局部广播问题是解决有着“瓶颈”的图上的全局广播问题的关键。
- 在现实意义下， k -局部广播问题的重要性在于，在例如社交网络中，每个节点通常只关心与其有着直接联系的或距离很短的节点的八卦。

2 算法与分析

2.1 1-局部广播问题

对于每个节点在每一轮中的行为，作者提出了如下算法。

算法 1: 1-DeterministicGossip

```

 $R = \{v\}$ 
 $i = 0$ 
while  $\Gamma(v) \setminus R \neq \emptyset$  do
     $i = i + 1$ 
    link to any new neighbor  $u_i \in \Gamma(v) \setminus R$ 
    PUSH: for  $j = i$  downto 1 do
        exchange rumors with  $u_j$ 
    end
    PULL: for  $j = 1$  to  $i$  do
        exchange rumors with  $u_j$ 
    end
    perform PULL, PUSH again
end

```

定理 2.1. 算法 1 以 $\log n$ 次迭代、不超过 $2(\log n + 1)^2$ 轮解决了 1-局部广播问题。

为了证明该定理，我们需要首先定义 i -tree。

定义 2.1. 一个 i -tree 为一个有根的、深度为 i 的、节点个数为 2^i 的树。其归纳构造如下：

1. 一个 0-tree 包含一个节点；
2. 对于 $i \in \mathbb{N}$ ， $(i+1)$ -tree 的构造是，取两个 i -tree，并将该二者的根节点相连，取其中一个的根节点为 $(i+1)$ -tree 的根节点。

现在，我们可以将 i -tree 和八卦的传播联系起来。假设在第 i 次迭代中， v_0 尚未结束八卦传播，那么我们可以构造其 i -tree τ_0 如下。 τ_0 的根为 v_0 ，其子节点依次为 u_1, \dots, u_{i-1} ，其分别第 $1, \dots, i-1$ 轮被 v_0 “发现”。我们递归地以每一个 u_j 为根节点，这里 $j \in [i-1]$ ，为其添加子节点 w_1, \dots, w_{j-1} ，这里新添加的子节点是前 j 次迭代之前 u_j 所“发现”的节点；现在我们对每个 u_j 的子节点递归地做类似的构造，如此反复。该构造的解释是， v_0

在第 j 次迭代仍然活跃的原因是, 其邻居 u_j 仍未被该节点所知, 而 u_j 在该次迭代活跃的原因是, 其邻居 v_0 仍未被该节点所知; 如此我们可以递归地解释该树中的每一个节点。

现在我们使用以下引理。

引理 2.2. 假设 u 和 v 在第 i 轮迭代开始时仍然活跃, τ_u 与 τ_v 是其相应的 i -tree。那么在该论中的 *PUSH* 操作后, 所有 τ_u 中的节点将会收到来自 u 的八卦 (对 v 与 τ_v 同理); 进一步, 如果 τ_u 和 τ_v 有公共节点, 那么在 *PUSH* 和 *PULL* 操作后, u, v 将会互相知道来自对方的八卦。

证明. 对树深度作归纳, 容易证明第一点。在此基础上, 假设 τ_u 和 τ_v 有公共节点 y , 那么根据上一单可知, 在 *PUSH* 操作中, y 将了解到 u, v 的八卦, 因此在 *PULL* 操作中, u, v 将获得来自 y 的对方的八卦。□

现在即可用该引理证明本节的定理。

对定理 2.1 的证明. 我们将证明使用算法 1 的迭代次数不超过 $\log n$ 。首先注意到该算法的对称性, 即当 v “发现” u 后, u 也“发现”了 v ——因为每轮中的 *PULL-PUSH* 操作和 *PUSH-PULL* 是对称的。现假设在第 i 轮后 u, v 互不知晓对方的八卦, 那么可以得到 τ_u 和 τ_v 没有公共节点。而在第 $\log n$ 次迭代后, τ_u 和 τ_v 的节点数分别为 n , 这与二者没有公共节点矛盾。因此我们可以得到结论 $\log n$ 次迭代可以保证 1-局部广播问题完成。通过计算可知, 解决该问题的总轮数不超过 $\sum_{i=1}^{\log n} 4i = 2 \log n (\log n + 1)$ 。□

2.2 k -局部广播问题

现在我们可以将上一小节的算法推广至 k -局部广播问题。首先可以直接将算法 1 运行 $k - 1$ 次——这显然能解决 k -局部广播问题。但考虑到, 在算法 1 中, 只有最后一步是保证了所有的邻居都收到了 v 信息, 而之前的每次迭代都是为了“发现”新节点而依次来扩张 τ_v 。因此只需要新增 $k - 1$ 次迭代, 每次迭代都向发现的邻居发送信息即可。因此, 算法的总运行轮数为 $2 \log n (k + \log n)$ 。如果我们想要解决全局广播问题, 只需要设置 $k = D$ 即可。

算法 2: k -DeterministicGossip

$R = \{v\}$

$i = 0$

Run 1-DeterministicGossip algorithm to discover neighbors **for** $s = 1$ **to** $k - 1$ **do**

PUSH: **for** $j = i$ *downto* 1 **do**

 exchange rumors with u_j

end

PULL: **for** $j = 1$ **to** i **do**

 exchange rumors with u_j

end

end

3 与随机 PULL-PUSH 模型的比较

在本节中，我们会生成一系列的图，并在此之上与 PULL-PUSH 模型比较全局广播问题的运行情况。（由于本文改进了 Censor-Hillel 等人提供的与电导率无关的随机算法 [Censor-Hillel et al., 2012]，且在思路上一脉相承，故不再与之比较。）

3.1 PULL-PUSH 模型

自从 Demers 等人介绍了 PULL-PUSH 模型 [Demers et al., 1987] 后，该模型得到了广泛的研究。该模型是对于八卦仅仅来源于单节点的情况所作的讨论：在每一轮中，持有该八卦的节点均匀地随机从其邻居中选择一个节点，并将该信息推送给它，而不持有该信息的节点也会均匀地随机从其邻居中选择一个节点，并向其请求信息。Chierichetti 等人分析了该模型的高效性，即只需要将经过 $O(\frac{\log n}{\phi})$ 轮传播，某节点的八卦即可以高概率被所有节点所知。这里，随机性扮演了重要的角色。

为适应我们的问题，即所有的节点都有其八卦的情况，我们将 PULL-PUSH 模型稍作修改。现我们定义每一轮中，持有**所有信息**（因为现在我们不仅仅考虑信息仅来自于一个节点）的节点将均匀地从其邻居中选择一个节点，将信息推送给它，而不持有所有信息的节点，将均匀地从其邻居中选择一个节点并将自己的信息与之**交换**（因为我们在算法 1 和 2 中总是和某个邻居交换信息）。

3.2 结果

我们将在连通的 $(2048, \frac{\ln 2048}{2048})$ -Erdős-Rényi 图 [Erdős and Rényi, 1960]、 $(1024, 12)$ -barbell 图 [Alon et al., 2007]、8-barbell 图（其中每个团为 K_{256} ）²与 2047-节点二叉树上比较 D-DeterministicGossip 与 PULL-PUSH 传播算法；其中，对于 Erdős-Rényi 模型我们生成多次，并且在每一个图上运行 PULL-PUSH 多次并取其平均。³

模型	直径	算法 2 轮数	PULL-PUSH 轮数	
			平均	最大
$(2048, \frac{\ln 2048}{2048})$ -Erdős-Rényi	7	16	11.4	12
$(1024, 12)$ -Barbell	14	17	31.8	35
8-Barbell(K_{256})	16	17	1137.4	1237
2047-节点二叉树	20	24	47.4	50

由此我们可以得出结论，虽然在高度连通的 Erdős-Rényi 图上 DeterministicGossip 算法比 PULL-PUSH 算法的运行轮数稍多，但可以看到在连通性不是那么强的图上，算法

²该图的电导率为 $\Theta(\frac{c}{n^2})$ [Censor-Hillel and Shachnai, 2011]，且通过计算发现其他图的电导率是该图的数十倍。

³由于电导率的计算是 NP-难的，因此本来试图利用 Cheeger 不等式 [Cheeger, 1969] 取 $\sqrt{2\lambda_2}$ 为电导率上界，这里 λ_2 是邻接矩阵的正则化 Laplacian 矩阵的第二小特征值。但是 SciPy 在计算特征值的时候，计算结果都极小且有的图上出现了复数，因此并没有将计算得到的特征值列在这里。

2 的运行轮数大约至多为 PULL-PUSH 算法轮数的一半；而且可以看到，有 Censor-Hillel 等人提出的由 K_{256} 组成的 8-Barbell 图上，确定型的算法 2 比随机的 PULL-PUSH 快了约 66 倍。

参考文献

- [Alon et al., 2007] Alon, N., Avin, C., Koucky, M., Kozma, G., Lotker, Z., and Tuttle, M. R. (2007). Many Random Walks Are Faster Than One. *arXiv:0705.0467 [math]*.
- [Censor-Hillel et al., 2012] Censor-Hillel, K., Haeupler, B., Kelner, J., and Maymounkov, P. (2012). Global Computation in a Poorly Connected World: Fast Rumor Spreading with No Dependence on Conductance. In *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing*, STOC '12, pages 961–970, New York, NY, USA. ACM.
- [Censor-Hillel and Shachnai, 2011] Censor-Hillel, K. and Shachnai, H. (2011). Fast Information Spreading in Graphs with Large Weak Conductance. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '11, pages 440–448, Philadelphia, PA, USA. Society for Industrial and Applied Mathematics.
- [Cheeger, 1969] Cheeger, J. (1969). A lower bound for the smallest eigenvalue of the Laplacian. *Proceedings of the Princeton conference in honor of Professor S. Bochner*, pages 195–199.
- [Chierichetti et al., 2018] Chierichetti, F., Giakkoupis, G., Lattanzi, S., and Panconesi, A. (2018). Rumor Spreading and Conductance. *Journal of the ACM*, 65(4):1–21.
- [Demers et al., 1987] Demers, A., Greene, D., Hauser, C., Irish, W., Larson, J., Shenker, S., Sturgis, H., Swinehart, D., and Terry, D. (1987). Epidemic Algorithms for Replicated Database Maintenance. In *Proceedings of the Sixth Annual ACM Symposium on Principles of Distributed Computing*, PODC '87, pages 1–12, New York, NY, USA. ACM.
- [Diestel, 2017] Diestel, R. (2017). *Graph Theory*. Graduate Texts in Mathematics. Springer-Verlag, Berlin Heidelberg, 5 edition.
- [Dunbar, 2004] Dunbar, R. I. M. (2004). Gossip in evolutionary perspective. *Review of General Psychology*, 8(2):100–110.
- [Erdős and Rényi, 1960] Erdős, P. and Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, 5:17–61.
- [Feinberg et al., 2012] Feinberg, M., Willer, R., Stellar, J., and Keltner, D. (2012). The virtues of gossip: Reputational information sharing as prosocial behavior. *Journal of Personality and Social Psychology*, 102(5):1015–1030.

- [Haeupler, 2015] Haeupler, B. (2015). Simple, Fast and Deterministic Gossip and Rumor Spreading. *J. ACM*, 62(6):47:1–47:18.
- [Jerrum and Sinclair, 1989] Jerrum, M. and Sinclair, A. (1989). Approximating the Permanent. *SIAM Journal on Computing*, 18(6):1149–1178.