

基于自适应随机行走的可扩展无偏抽样方法^{*}

符永铨¹⁺, 王意洁¹, 周婧²

¹(国防科学技术大学 计算机学院 并行与分布处理国家重点实验室,湖南 长沙 410073)

²(工程兵指挥学院 计算机教研室,江苏 徐州 221000)

A Scalable Unbiased Sampling Method Based on Multi-Peer Adaptive Random Walk

FU Yong-Quan¹⁺, WANG Yi-Jie¹, ZHOU Jing²

¹(National Key Laboratory for Parallel and Distributed Processing, School of Computer, National University of Defense Technology, Changsha 410073, China)

²(Staff Room of Computer, Command Academy of the Corps of Engineers, Xuzhou 221000, China)

+ Corresponding author: E-mail: quanyongf@gmail.com

Fu YQ, Wang YJ, Zhou J. A scalable unbiased sampling method based on multi-peer adaptive random walk.

Journal of Software, 2009,20(3):630–643. <http://www.jos.org.cn/1000-9825/3206.htm>

Abstract: To deal with the scalable and fast unbiased sampling problems in unstructured P2P systems, a sampling method based on multi-peer adaptive random walk (SMARW) is proposed. In the method, based on the multi-peer random walk process, a set of provisional peers are selected as agents which start the sampling processes, by which the sampling process is speeded up with receiving a set of tunable number samples each time; Meanwhile, after receiving new samples earlier agents are replaced with these new samples which repeat the sampling process. With this simple replacement, it can be guaranteed with high probability that the system can reach the optimal load balance; furthermore, SMARW adopts an adaptive distributed random walk adjustment process to increase the convergence rate of the sampling process. A detailed theoretical analysis and performance evaluation confirm that SMARW has a high level of unbiased sampling and near-optimal load balancing capability.

Key words: unbiased sampling; unstructured P2P system; randomized algorithm; scalability

摘要: 针对非结构化P2P系统中可扩展的快速无偏抽样问题,提出了一种基于多个peer自适应随机行走的抽样方法SMARW.在该方法中,基于代理随机行走选择一组临时的peer执行抽样过程,一次产生一组可调数目的抽样节点,提高了抽样速度,选择每次产生的抽样节点作为临时peer进行新的抽样过程,这种简单的方法可以保证系统具有近似最优的系统负载均衡程度.同时,SMARW利用自适应的分布式随机行走修正过程提高抽样过程的收敛速度.理

* Supported by the National Natural Science Foundation of China under Grant Nos.60873215, 60621003 (国家自然科学基金); the National Basic Research Program of China under Grant No.2005CB321801 (国家重点基础研究发展计划(973)); the Specialized Research Fund for the Doctoral Program of Higher Education of China No.200899980003 (高等学校博士学科点专项科研基金); the Foundation for the Author of National Excellent Doctoral Dissertation of China under Grant No.200141 (高等学校全国优秀博士学位论文作者专项); the Graduate Research Innovation Fund of NUDT of China under Grant No.B080602 (国防科学技术大学优秀研究生创新资助)

Received 2007-06-30; Accepted 2007-10-12

论分析和模拟测试表明,SMARW 方法具有较高的无偏抽样能力以及近似最优的系统负载均衡程度。

关键词: 无偏抽样;非结构化 P2P 系统;随机算法;可扩展性

中图法分类号: TP393 **文献标识码:** A

随着 P2P 网络研究的不断深入,许多研究者认识到在较大规模且强网络波动的 P2P(peer-to-peer)系统中,设计基于维护少量系统状态且保证性能的随机算法更为有效.无偏的抽样技术通过均匀随机选择系统中在线 peer 作为随机算法的输入,为大量随机算法提供了基础性的功能(本文不考虑 P2P 系统中带有偏好的抽样技术).非结构化 P2P 系统中典型的随机算法,如构建 P2P 网络^[1,2],资源查找^[3],数据分发^[4-6],研究 P2P 网络宏观特征等^[7-10];结构化 P2P 系统中路由路径优化^[11,12].然而,上述随机算法的性能取决于无偏抽样过程的有效性,高效的抽样过程必须满足以下条件:

- 无偏的:最基本要求,抽样过程必须是无偏好的,即选择任意节点的概率相同.
- 快速收敛的:抽样过程必须具有较低的等待延迟.
- 可扩展的:抽样过程必须具有较高的负载均衡程度,避免出现网络通信热点(hotspot).

考虑在非结构化 P2P 系统可扩展的快速无偏抽样.原因是非结构化 P2P 系统中基于随机化邻居抽样而能够适应更广泛的 P2P 系统,在抽样过程中随着某些节点抽样数目的增大,该发起者周围可能会引发负载不均衡问题.这种问题在抽样节点数目较大的应用中尤为突出.在这些应用中,抽样过程的负载均衡程度直接影响了系统的正常工作性能.因此,研究抽样过程的可扩展性是一个非常紧迫的问题.

已有的非结构化 P2P 系统抽样方法研究主要关注于无偏性和收敛速度,没有考虑在大规模抽样下的扩展性问题.已有的针对非结构化 P2P 系统的抽样方法主要基于随机行走(第 2.1 节对随机行走进行介绍,Awan 等人^[13]考虑在静态非结构化 P2P 网络中提高抽样过程的收敛速度,从而提高了抽样的无偏程度,然而,抽样方法执行次数较多,依赖于多个静态参数,导致缺乏在实际网络环境下的自适应能力.Stutzbach 等人^[14]首次利用 Metropolis Hastings 方法^[15]在动态的 P2P 网络环境下进行抽样;然而在模拟和真实环境下的实验结果差别非常大,同时没有考虑提高抽样过程的收敛速度.上述工作^[13,14]均没有考虑抽样过程的可扩展性.

本文的目标是在满足基于随机行走抽样快速收敛的基础上,消除任何可能出现的系统负载不均衡影响.提出了一种基于多个 peer 自适应随机行走进行可扩展的无偏抽样方法(sampling by multi-peer adaptive random walk,简称 SMARW).在 SMARW 中,利用代理随机行走(multi-peer random walk,简称 MPRW)过程一次产生一组可调数目的抽样节点,提高了抽样速度,隐藏了单个随机行走下的固有等待延迟,同时,选择每次产生的抽样节点作为临时 peer 进行新的抽样过程,在理论上证明这种简单的方法可以保证系统具有近似最优的负载均衡程度;为提高 MPRW 抽样过程的无偏能力,提出了自适应的分布式随机行走修正过程,该过程分布式的提高随机行走收敛到稳态分布的速度,同时不需要设置静态的参数且在 $O(\log N)$ 轮数内完成,具有较高的自适应能力.理论分析和模拟测试证明,SMARW 方法具有较高的无偏抽样水平以及高度的可扩展能力.

1 非结构化 P2P 网络无偏抽样

为便于分析,用无向图 $G=(V,E)$ 表示一个抽象的非结构化 P2P 网络模型,无向图中的节点代表 P2P 网络中的节点,图中节点间的无向边表示 P2P 网络节点间的双向通信链接.系统模型是静态的,不考虑系统中可能的网络波动场景.

1.1 无偏抽样相关概念

首先给出无偏抽样的确切含义.

定义 1(无偏抽样). 一种算法选择一个连通的网络 $G=(V,E)$ 中任意节点的概率相互独立且为 $1/N$,其中 $N=|V|$,则称被选择的节点为无偏抽样.

随机行走提供了分布式的抽样方法:网络中某个节点自主地发起随机行走消息,并基于概率的方式确定下

一跳消息转发节点,消息经过一定数量的中间转发节点后终止于一个目的节点,该目的节点即为新的抽样.随机行走终止于任意节点的概率确定了目的节点是否为无偏抽样.

定义 2(无偏抽样随机行走). 若一个随机行走终止于某个连通的网络 $G=(V,E)$ 中任意节点 v 的概率为 $1/N$, 则称该行走为一个无偏抽样随机行走.

在静态的网络图中随机行走可以形式化地定义为一个有限的同构马尔可夫链,其状态集为 V ,边集为 E ,转移矩阵 $P=(P_{ij})$,代表任意节点 i 转移到节点 j 的概率, $(i,j) \in V \times V, \sum_j P_{ij} = 1, i \in [1, N]$. 例如,最简单的转移矩阵(记为 Naive): $p_{ij} = 1/\text{deg}(i), (i,j) \in E, \text{deg}(i)$ 为节点 i 的连接度数.非结构化 P2P 网络上的随机行走可以被视为一个有限的同构马尔可夫链^[13,14].若概率分布向量 v 描述随机行走每个节点的概率分布,即 $\sum_i v(i) = 1$,其中 $v(i)$ 为向量 v 的第 i 项,则 $v' = vP$ 描述在下一跳步时的概率分布,类似地, vP^r 为在 r 跳步后的概率分布.

对于一个有限的同构马尔可夫链,随机行走停留在任意节点 x 的概率收敛到一个稳态分布 π ,该稳态分布有且唯一,且下式成立^[16]: $\pi(x) \times P_{xy} = \pi(y) \times P_{yx}$.称达到稳态下的概率分布向量 v 为稳态分布向量.如果每隔 r 步选择 1 个节点作为抽样,对于足够大的 r :

- r 跳步后的概率分布向量与起始的概率分布向量相互独立.
- 稳态分布 $\pi(x)$ 预先可知,并且 $\pi(x)$ 偏向于连接度高的节点,即 $\pi(x)$ 不是无偏的.

在实际的网络环境下,对于具有一个强随机分量的图(如非结构化 P2P 网络),随机行走接近稳态分布的跳步数仅需要大于网络中所有节点数目的对数,即 $r \geq O(\log|V|)$ ^[12-14],称该跳步数 r 为随机行走的固有等待时间.这里,给出随机行走无偏稳态分布的充要条件:

引理 1(无偏稳态分布充要条件^[13]). 设 M 为一个有限的同构马尔可夫链, M 的稳态分布 $\pi(x)$ 为无偏的当且仅当 M 的转移矩阵为对称的.

证明:由 $\pi(x) \times P_{xy} = \pi(y) \times P_{yx}$ 易得. □

引理 1 提供了实现无偏稳态分布的一般方法,随机行走只需要保证转移矩阵为对称的.典型的如 Metropolis-Hastings 方法^[15],其转移矩阵 Q 为

$$Q(x, y) = \begin{cases} \frac{1}{\max(\text{deg}(x), \text{deg}(y))}, & x \neq y \\ 1 - \sum_{x \neq y} Q(x, y), & x = y \end{cases}$$

显然, Metropolis-Hastings 转移矩阵为对称的,因此 Metropolis-Hastings 方法产生无偏的稳态分布,然而,该方法具有较高的自转移概率 $Q(x, x)$,而自转移概率越高,行走的收敛速度越低,证明过程见文献[13].尤其是在是在系统中存在一些节点连接度较小而其邻居节点连接度较大的现象时,随机行走的低收敛速度问题更加突出.

1.2 基于随机行走无偏抽样的可扩展性

为提高随机行走的快速收敛能力,降低抽样过程的等待延迟问题,一方面可以通过降低单个随机行走的自转移概率以提高达到稳态分布的速度^[13];另一方面,可以通过执行并行的多个随机行走来隐藏单个随机行走的等待延迟^[3].前者无法降低一个随机行走的固有等待时间;后者由于采用并行随机行走的方式更具有灵活性,然而,后者从一个节点通过邻居节点发出大量的随机行走消息,可能导致在非结构化 P2P 网络中出现网络通信热点.

假定非结构化 P2P 网络中一个节点 u (称为发起节点)基于随机行走进行抽样,随机行走消息均通过邻居节点进行逐跳地转发,每次随机行走的固有等待时间为 r 跳步^[3,13,14].定理 1 和定理 2 证明了在上述假定下,即使相邻的两个节点,转发随机行走消息网络通信负载差别也可能很大.进一步地,推论 1 显示,在特殊的 d 规则图下转发随机行走消息的网络通信负载沿距发起节点的跳步数呈指数级降低,即产生了严重的系统负载不均衡问题.

考虑参与转发行走消息的节点子集,称一个节点 w 距发起节点 u 为 m 跳步当且仅当节点 w 距离发起节点 u 的最短跳步数为 m .

定理 1(Naive 相邻节点消息转发差别). 假定系统采用 Naive 随机行走方法进行抽样,考察在任意的拓扑图

中距离发起节点为 m 跳步和 $m+1$ 跳步的相邻节点 v_m 和 v_{m+1} ,发起节点在利用邻居节点发送一定数量的随机行走消息后, v_m 与 v_{m+1} 消息转发数目比例期望值下限为 $\Omega(1/d_{m+1}^2)$,上限为 $O(d_m)$.其中, d_m, d_{m+1} 分别为系统中节点 v_m, v_{m+1} 的连接度.

证明:利用宽度优先遍历策略(BFS)沿发起节点出发记录拓扑图中节点,获得一个按距发起节点跳步数递增顺序的节点排序: $v_0, v_{11}, v_{12}, \dots, v_{1m}, v_{21}, v_{22}, \dots, v_0$ 为发起节点, v_{ij} 为距 v_0 跳步数为 i 的第 j 个节点.如图 1 中设 A 为发起节点,按照 BFS 排序得 ABCDFE.

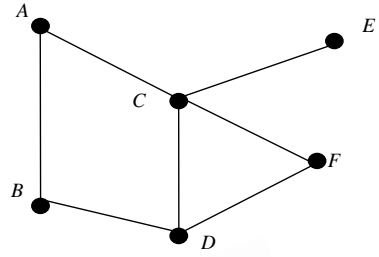


Fig.1 An unstructured P2P network

图 1 一个非结构化 P2P 网络模型

- (1) 按 Naive 随机行走规则,距发起节点 m 跳步的转发节点 v_m 均匀、随机地选择某个邻居节点 w 传播消息(w 距发起节点跳步数最多为 $m+1$),则在 v_m 接收并转发 k 个消息的情况下,节点 v_{m+1} 自节点 v_m 获取的消息数期望值为 $k/d_i, d_i$ 为节点 i 连接度,故 v_m 和 v_{m+1} 消息转发数目比例期望值上限 $\leq k/(k/d_m) = d_m$,即 $O(d_m)$.

- (2) 考虑节点 v_{m+1}, v_{m+1} 可能从距发起节点跳步 $\geq m$ 的邻居接收额外行走消息. v_{m+1} 获取额外消息期望值 S_e :

$$S_e \leq \sum_{i=1}^{d_{m+1}-1} S_i/d_i \leq \sum_{i=1}^{d_{m+1}-1} S^* = (d-1) \times S^* .$$

S_i 为节点 v_{m+1} 邻居节点 i 接收的行走消息数目, d_i 为节点 i 的连接度,且

$$S^* = \max(S_i/d_i) = S'/d' .$$

设 S'' 为 v_{m+1} 节点转发消息数目, S, d_m 分别为节点 v_m 转发消息数和连接度, S', d' 分别为 S^* 对应节点的转发消息数和连接度,显然,该节点距发起节点的跳步数至少为 m .同时有如下不等关系成立:

$$\frac{S''}{d_{m+1}} \leq S, \frac{S'}{d'} \leq S'', \frac{S''}{d_{m+1}} \leq \frac{S'}{d'}, \frac{S}{d_m} \leq \frac{S'}{d'} .$$

系统中距离发起节点为 m 和 $m+1$ 跳步的相邻节点 v_m, v_{m+1} 消息转发的比例下限为

$$\begin{aligned} \frac{S}{d_m + S_e} &\geq \frac{S}{d_m + (d_{m+1}-1) \times S^*} = \frac{S}{d_m + (d_{m+1}-1) \times \frac{S'}{d'}} \geq \frac{S}{\frac{S'}{d'} + \frac{d_{m+1}-1}{d'} \times S'} = \frac{d'}{d_{m+1}} \times \frac{S}{S'} \geq \\ &\frac{1}{d_{m+1}} \times \frac{d'}{d_{m+1}} \times \frac{S''}{S'} \left(\text{由 } \frac{S''}{d_{m+1}} \leq S \right) \geq \frac{1}{d_{m+1}} \times \frac{d'}{d_{m+1}} \times \frac{1}{d'} \left(\text{由 } \frac{S'}{d'} \leq S'' \right) = \Omega\left(\frac{1}{d_{m+1}^2}\right) . \end{aligned}$$

定理 1 证毕. \square

定理 2(Metropolis-Hastings 相邻节点消息转发差别). 考察任意拓扑图中距发起节点 $m, m+1$ 跳步的相邻节点 v_m, v_{m+1} ,基于 Metropolis-Hastings 抽样的情况下,节点 v_m 和 v_{m+1} 消息转发数目比例期望值下限为 $\Omega(1/\max(d_m, d_{m+1})^2)$,上限为 $O(\max(d_m, d_{m+1}))$. d_m, d_{m+1} 分别为系统中节点 v_m, v_{m+1} 的连接度.

证明:与定理 1 的证明方式类似,由第 2.1 节 Metropolis-Hastings 方法,任意节点 v_m 转移到逻辑邻居 v_{m+1} 的概率为 $1/\max(d_m, d_{m+1})$,按照定理 1 方式将节点排序,选择每个距发起节点 m 跳步的转发节点 v_m ,

- (1) 在 v_m 转发 k 个消息下, v_{m+1} 自节点 v_m 获取的消息期望值为 $k/\max(d_m, d_{m+1})$,故上限 $O(\max(d_m, d_{m+1}))$;
- (2) v_{m+1} 获取的额外行走消息期望值 S_e :

$$S_e \leq \sum_{i=1}^{d_{m+1}-1} S_i/\max(d_m, d_{m+1}) \leq \sum_{i=1}^{d_{m+1}-1} S^* = (d_{m+1}-1) \cdot S^* , \text{其中 } S^* = \max(S_i/\max(d_m, d_{m+1})) = \frac{S'}{d'} ,$$

且如下不等式成立:

$$\frac{S''}{\max(d_{m+1}, d_m)} \leq S, \frac{S'}{\max(d', d_{m+1})} \leq S'', \frac{S''}{\max(d_{m+1}, d_m)} \leq \frac{S'}{\max(d', d_{m+1})}, \frac{S}{\max(d_{m+1}, d_m)} \leq \frac{S'}{\max(d', d_{m+1})} .$$

按定理 1 步骤(2)的同样方式可得下限为 $\Omega(1/\max(d_m, d_{m+1})^2)$.

图 2 给出了定理 1 和定理 2 接近上限和下限的两类 P2P 网络局部的拓扑关系.新的随机行走消息流向按照距发起节点的跳步数递增的顺序转发,图 2 中新的消息从 v_m 和 v_k 向下流动.图 2(a)中节点 v_m, v_{m+2} 分别距离发起节点为 m 和 $m+2$ 跳步,图 2(a)接近上限, v_m 的随机行走消息分散到各个分支, v_m 的转发消息数量约为 v_{m+2} 的 4 倍,在具有连接度上限的 P2P 网络下较为普遍,如文献[1];图 2(b)中节点 v_{k+1}, v_{k+2} 分别距发起节点 $k+1, k+2$ 跳步,图 2(b)接近下限, v_k 的消息在分散后又会聚到 v_{k+2}, v_{k+1} 的转发消息数量约为 v_{k+2} 1/3 倍,在满足幂率分布下的 P2P 网络较为常见,如 Gnutella^[8].

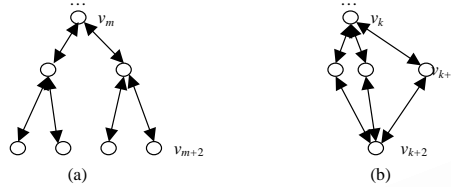


Fig.2 Forwarding process of random walk messages

图 2 随机行走消息转发示意图

以下推论给出了在特殊 d -规则图下消息转发数目与距发起节点跳步数的关系.

推论 1. d -规则图(节点连接度均为 d)中,若距发起节点为 k 跳步的每个节点均有不相交的 $d-1$ 个新邻居节点,且新邻居节点距发起节点为 $k+1$ 跳步,则在 Naïve 或 Metropolis-Hastings 随机行走下,随着距发起节点跳步数的增加,参与消息转发的节点转发的消息数期望值按照与发起节点的距离 k 呈 $O(d^k)$ 指数级降低.

推论 1 对应应在图 2(a)拓扑结构,行走消息负载自发起节点开始沿着分支向下呈指数级降低.为衡量抽样方法的可扩展性,在一个节点发起一组随机行走消息后,将系统中负载均衡被破坏的程度量化为负载失衡度.

定义 3(负载失衡度). 将所有节点转发的随机行走消息数按升序排列为 $S = \{s_1, s_2, \dots, s_N\}$, $\forall i, j \in [1, N], i < j, s_i \leq s_j$, 设 $Msg_{\max} = s_N, Msg_{\text{median}} = s_{\lceil (1+N)/2 \rceil}$. 系统中负载失衡度 LIF(load imbalance factor)定义为 $LIF = Msg_{\max} / Msg_{\text{median}}$.

第 4 节利用模拟测试来计算不同行走方法的 LIF. LIF 反映了热点地区的最坏情况, LIF 越低,系统的负载均衡程度越高.理想的最优系统负载均衡情况下,每个节点转发随机行走的消息数量不存在差别,即 LIF 值为 1.

2 基于多个 peer 自适应随机行走的可扩展无偏抽样方法

2.1 代理随机行走

在代理随机行走(MPRW)中,某 peer initiator(称为发起节点)随机选择一组临时的 peer 作为代理后,各个代理独立地发送随机行走消息,最终,消息停止的节点作为新的抽样节点;该抽样节点发送响应消息至原发起节点 initiator,同时,该次行走的原代理被撤销,该抽样节点被任命为新的代理节点继续进行随机行走.

代理随机行走(MPRW)包括初始化代理节点、代理节点进行随机行走以及代理节点更新 3 个过程.

(1) 初始化代理节点

initiator 构造一个类型为 GetAgent 的消息,设定代理节点数目 K ,然后均匀地随机选择一个邻居节点进行转发;邻居节点将本节点联系信息(IP 地址,端口)填入 GetAgent 消息,然后,按照上述方式继续转发该消息.转发过程直至 GetAgent 消息中无重复的节点联系信息数目为 K 时结束,GetAgent 消息被转发回 initiator.

(2) 代理节点进行随机行走

initiator 确定需要的抽样节点数目 I 作为抽样过程的终止条件,设定每个行走消息的最大跳步数 TTL ,然后 initiator 发送行走消息至每个代理节点.每个代理节点独立地利用基于 DAM 过程设定转移概率的

`getNextNeighbor()` 确定下一跳步邻居节点 w , 然后转发行走消息至节点 w , 之后该行走消息均按照上述方式转发. 在每个随机行走消息达到最大跳步数 TTL 时, 当前行走消息所在的节点 n 作为新的抽样点发送响应消息到发起节点 $initiator$.

(2) 代理节点更新

在一次代理节点随机行走过程结束后, 发起本次行走消息的代理被取消, $initiator$ 设定本次随机行走的抽样节点 n 作为新的代理节点, 并构造新的行走消息发送到代理节点 n . 代理节点 n 继续执行过程(2). $initiator$ 在抽样节点数目达到所需值 I 后取消所有的代理节点, 终止随机行走过程.

关于代理随机行走的说明如下:

- 代理随机行走与并行的多个随机行走^[3]的关键区别在于, MPRW 利用抽样节点进行新的随机行走, 降低了邻居节点的转发消息开销.

- 发起者确定代理节点数目, 代理节点数目越高, 则发起者获取一个抽样的等待时间延迟越低. 假定一个代理节点完成随机行走的平均时间为 T , 代理节点数目为 k , 则发起者平均等待 T/k 间隔即获得一个抽样节点.

- 每次行走消息的 TTL 需要设定为大于 $\log N$, 以接近随机行走的稳态分布, 其中 N 为系统中节点数量. 在实际网络环境下, N 可能无法精确判定, 一般满足 $TTL > 40$ 即可.

- 由第 3.1 节稳态分布的性质可知, 初始化代理节点的选择不影响随机行走的最终抽样节点, 因此, 发起节点可以利用缓存的历史节点信息作为初始化的代理节点, 以消除初始化代理节点的等待延迟.

2.2 自适应的分布式随机行走修正过程

针对 Metropolis-Hastings 方法引入自转移概率导致收敛速度降低的问题^[7], 提出了自适应的降低系统中 peer 自转移概率的分布式随机行走修正过程(dynamic-adjusting metropolis hastings, 简称 DAM).

DAM 是一个分布式的降低自转移概率的过程, 它包括两个子过程:(1) 初始化, 利用 Metropolis-Hastings 方法初始化系统的转移概率矩阵 P ; (2) 动态调整, 通过多轮的方式更新转移概率矩阵 P . 在每一轮, 任意节点 u 如果当前自转移概率 > 0 且存在自转移概率 (> 0) 的邻居节点, 则节点 u 触发动态调整过程: 节点 u 选择一个最小自转移概率 (> 0) 的邻居节点 y , 然后节点 u 和 y 更新自己的自转移概率. 具体过程如下:

(1) 初始化: 每个节点 x 初始化转移概率向量是

$$p^0(x, y) = \begin{cases} \frac{1}{\max(\deg(x), \deg(y))}, & x \neq y \wedge y \in x.\text{neighbor} \\ 1 - \sum_{x \neq y} p^0(x, y), & x = y \end{cases}$$

$x.\text{neighbor}$ 为 x 的邻居节点集合.

(2) 降低自转移概率: 在第 i 轮, 如果任意节点 x 当前自转移概率 > 0 , 节点 x 选择一个具有最小自转移概率 (> 0) 的邻居节点 y_0 , 然后, 节点 x 和 y_0 更新自己的自转移概率; 如果不存在该类型邻居, 节点 x 结束 DAM 过程. 节点 x 和 y_0 转移概率向量更新如下:

$$P^i(x, y) = \begin{cases} P^{i-1}(x, y_0) + \rho, & \forall u \in x.\text{neighbor} (p_{uu}^{i-1} \geq p_{y_0 y_0}^{i-1} \wedge p_{y_0 y_0}^{i-1} > 0) \\ P^{i-1}(x, y), & y \neq y_0 \wedge y \neq x \\ 1 - \sum_{x \neq y} P^i(x, y), & x = y \end{cases},$$

$$P^i(y_0, y) = \begin{cases} P^{i-1}(y_0, x) + \rho, & y = x \\ P^{i-1}(y_0, y), & y \neq y_0 \wedge y \neq x, \\ 1 - \sum_{y \neq y_0} P^i(y_0, y), & y = y_0 \end{cases}$$

$$\rho = \min(P^{i-1}(x, x), P^{i-1}(y_0, y_0)).$$

利用 DAM 方法确定节点的转移概率向量后, 获得当前节点的下一跳步转发节点过程:

$n.\text{getNextNeighbor}(\text{Vector } \text{vec})$

```

//vec 为转移到每个邻居节点的概率,vec 最后一个元素为自转移概率  $p_m$ ,vec 的索引值  $i$  对应第  $i$  个邻居
p ← generate a number uniform randomly from [0,1]
foreach  $i \in \text{vec}$ 
  if  $p \leq \text{vec.get}(i)$  //vec.get(i)返回本节点到第  $i$  个邻居节点的转移概率
    return  $n.\text{getNeighbor}(i)$  //获取第  $i$  个邻居,vec 的最后一个节点为本节点
  else  $p = p - \text{vec.get}(i)$ 

```

DAM 过程的特点包括:(1) 通过一步缩减自转移概率为 0,极大地缩短了执行轮数,降低了修正过程的时间开销;(2) 不需要配置静态参数,具有较高的自适应能力;(3) 任意节点在每次修正过程中均选择邻居节点中最小的自转移概率作为修正幅度,降低了转移到某个邻居节点概率过大的可能。

适应动态网络波动:为了适应网络波动,SMARW 方法中 DAM 过程修改如下:

- 若任意节点 i 检测到邻居节点 j 失效,清除到节点 j 转移概率,并设置节点 i 自转移概率 $p'(i,i)$ 修改为

$$p'(i,i) = p'(i,i) + p'(i,j);$$

- 若节点 i 增加新的邻居节点 j ,则节点 i,j 分别设置转移概率 $p'(i,j), p'(j,i)$ 为 Metropolis-Hastings 方法下的初始概率 $p'(i,j) = p'(j,i) = 1/\max(\text{deg}(i), \text{deg}(j))$,并分别减小自转移概率 $p'(i,i)$ 以保持对称性.如当前自转移概率 $p'(i,i)$ 或 $p'(j,j) < p'(i,j)$,则 $p'(i,i)$ 或 $p'(j,j)$ 回滚(rollback)到最后一次 DAM 过程前的状态(假设在 $(t-1)$ 轮与节点 k 降低自转移概率),并更新到对应节点 k 的转移概率。

上述过程保证任意节点转移概率矩阵为对称的,由第 2.1 节引理 1 可知,DAM 过程具有无偏稳态分布.上述动态调整结束后,节点 i 继续降低自转移概率。

2.3 SMARW理论分析

1) SMARW 收敛速度与无偏性分析

定理 3. DAM 方法执行的轮数上限为 $\log N, N$ 为系统中的节点数量。

证明:考察 DAM 方法的每一轮执行过程.在第 i 轮更新过程中,设存在 N' 数目的节点集合 S^i 参与降低自转移概率过程($N' \leq N$).选择 S^i 中任意的待更新自转移概率的节点对 (x, y_0) ,其更新幅度 ρ 为

$$\rho = \min(P^{i-1}(x, x), P^{i-1}(y_0, y_0)).$$

因此,该轮结束后,节点 x 或 y_0 的自转移概率被缩减为 0,即至少其中一个节点不再参与 DAM 过程.从而在第 i 轮更新结束后,系统中仍然需要执行 DAM 过程的节点数目 $\leq |S^i|/2$.因此,在 k 轮 DAM 过程后,需要更新自转移概率的节点数目 $\leq 2^{-k} \cdot N$.故 DAM 方法执行的轮数上限为 $\log N$.定理 3 得证. \square

定理 3 证实 DAM 过程的执行轮数以系统中节点数目的对数为上限,能够快速地消除自转移概率带来的收敛速度降低的影响,因而 SMARW 具有较高的收敛速度.定理 4 保证基于 SMARW 方法的抽样为无偏的。

定理 4. 设 M 为一个有限的同构马尔可夫链,基于 SMARW 方法下 M 的稳态分布为无偏的。

证明:由定理 3 可知,DAM 方法的转移概率矩阵更新的执行轮数上限为 $\log N$.考察 DAM 下的每一轮的更新过程.利用归纳法证明:

(1) 在 $i=0$ 轮时, M 的转移概率矩阵 P 利用 Metropolis-Hastings 方法进行初始化,显然 P 为对称矩阵.由引理 1 可知,在转移概率矩阵 P 下 M 的稳态分布为无偏的。

(2) 假设在 $i \leq k$ 轮时 M 的稳态分布为无偏的.显然,转移概率矩阵 P^i 为对称矩阵.在 $i=k+1$ 时,设节点 x_0 随机选择一个邻居 y_0 ,然后基于 DAM 方法互相更新转移概率.由假设可得

$$P^i(x, y) = P^i(y, x), \forall x, y \in N.$$

并且由更新过程,下述等式成立:

$$P^{i+1}(x_0, y_0) = P^i(x_0, y_0) + \rho.$$

$$P^{i+1}(y_0, x_0) = P^i(y_0, x_0) + \rho.$$

$$P^{i+1}(x, y) = P^i(x, y), \forall x, y \notin \{x_0, y_0\} \wedge x \in N, y \in N.$$

因此, $P^{i+1}(x, y) = P^{i+1}(y, x), \forall x, y \in N$.

故 M 的转移概率矩阵 P^{i+1} 保持为对称矩阵, 由引理 1 可知, 在 $i=k+1$ 时, M 稳态分布保持为无偏的.

综合(1)、(2), 定理 4 得证. \square

由定理 4 可知, SMARW 过程保证随机行走获得的抽样为无偏的, 因此, 利用 `getNextNeighbor()` 方法获取下一跳步节点的随机行走能够保证抽样过程无偏. 下面分析 SMARW 的可扩展性.

2) SMARW 可扩展性分析

定理 5. 设系统规模为 N , 考察一个发起节点 u 开始的代理随机行走过程, 设任意代理节点每次随机行走的长度为 L , 代理节点执行随机行走的总次数为 m , 则任意 peer 被访问次数的期望值为 mL/K .

证明: 假设发起节点选择 K 个代理, 每个代理相互独立. 第 3.1 节初始化代理节点过程选择的代理节点并不严格相互独立, 然而在每个代理执行完一次随机行走后, 由行走过程达到稳态分布可知, 每个新的代理节点完全相互独立. 因此可以认为, 在整个行走过程中, 代理节点之间相互独立.

(1) 自代理节点第 1 次进行随机行走过程开始, 以最初的 K 个代理节点分别作为起始点, 以结束行走时对应的 K 个新代理节点为终节点, 构造 K 个独立的代理节点生成序列:

$$\{\{agent_{i1}, agent_{i2}, \dots, agent_{in}\}, i \in [1, K]\},$$

其中, 每个序列 $\{agent_{i1}, agent_{i2}, \dots, agent_{in}\}$ 中首元素 $agent_{i1}$ 为第 i 个初始化代理节点, $agent_{ij}$ 为以 $agent_{i1}$ 为初始化代理节点且进行 $j-1$ 次行走后对应的代理节点, 而 $agent_{in}$ 为以 $agent_{i1}$ 为初始化代理节点且进行 $n-1$ 次行走后对应的代理节点, n 为常数. 在下面分析中, 假定每个序列长度相同, 长度不同的情形分析过程相同.

(2) 上述 K 个代理节点序列将所有的行走过程分割为 K 个独立的行走过程: 每个过程的起始节点为 1 个初始化代理节点 $agent_{i1}, i \in [1, K]$; 该行走过程的终节点为 $agent_{in}$, 且每个行走的总长度为 mL/K . 考察 K 个独立的行走过程中每个节点 u 被访问的次数, 首先, 有限同构马尔可夫链中两次访问同一节点的访问时间(access time) $H_{ii} = 1/\pi(u)$ ^[3], 其次, 由代理随机行走的转移概率矩阵为对称矩阵得知, 每个代理节点的随机行走为无偏的, 即 $\pi(u) = 1/N$, 故在每个独立的行走过程节点 i 被访问次数的期望值为 $(mL/K)/H_{ii} = (mL/K)/N$. 由不同代理之间的独立性可知, 在所有的 K 个行走过程中, 每个节点被访问次数期望值为

$$\left(\frac{mL}{K}\right)/N \cdot K = mL/N.$$

定理 5 证毕. \square

定理 5 给出了代理随机行走下每个节点转发行走消息的期望负载, 上述结论与代理节点数目无关, 只与代理节点进行行走的总次数和每次行走的跳步数有关. 以下推论显然成立:

推论 2(SMARW 相邻节点消息转发差别). SMARW 方法下任意相邻节点转发消息数目比例期望值为 1.

推论 2 说明 SMARW 方法能够确保系统中任意节点网络通信开销期望值相同. 定理 6 给出了每个节点网络通信开销偏离期望值的程度.

定理 6. 设系统规模为 N , 考察一个发起节点 u 开始的代理随机行走过程, 设任意代理节点每次随机行走的长度为 L , 代理节点执行随机行走的总次数为 m , 则代理随机行走中任意节点转发行走消息的数量 X 以极大概率

$(1-O(N^{-C}))$ 为 $X \in \frac{mL}{N} \pm O\left(\sqrt{\frac{mL}{N} \log N}\right)$, 其中 C 为任意正实数.

证明: 设定随机变量 X 表示任意节点 u 转发代理随机行走消息数量, 由定理 5 可知, $E(X) = mL/N$, 记 $\mu = E(X)$. 现在利用 Chernoff 范围公式推导 X 的偏差.

$$\text{由 } P(X \geq \mu + \lambda) = P\left(X \geq \mu \left(1 + \frac{\lambda}{\mu}\right)\right) \leq e^{-\left(\frac{\lambda}{\mu}\right)^2 \frac{\mu}{3}} = e^{-\frac{\lambda^2}{3\mu}} = e^{-\frac{\lambda^2 N}{3mL}}, \text{ 设定 } \lambda = O\left(\sqrt{\frac{mL}{N} \log N}\right), \text{ 则}$$

$$P\left(X \geq \mu \left(1 + \frac{\lambda}{\mu}\right)\right) \leq e^{-\frac{1}{3} \log N} = \frac{1}{N^C}, C = \frac{1}{3}.$$

因此,以较高概率($1-O(N^{-c})$),任意节点 μ 转发代理随机行走消息数量为 $X \in \frac{mL}{N} \pm O\left(\sqrt{\frac{mL}{N} \log N}\right)$.

定理 6 证毕. □

定理 6 给出了每个节点转发行走消息数目的波动范围,在 $m \leq O\left(\frac{N}{\log N}\right)$ 下,

$$\sqrt{\frac{mL}{N} \log N} \rightarrow O(1).$$

因此,SMARW 具有较高的负载均衡程度.综合推论 2 和定理 6,大量抽样情况下 SMARW 方法能够以极大概率保证系统达到近似最优的系统负载均衡程度.而第 2.2 节 Naïve 和 Metropolis-Hastings 方法,即使相邻节点其转发的随机行走消息数目差别也可能很大,因而,Naïve 和 Metropolis-Hastings 方法无法保证系统的负载均衡程度.

3 模拟测试

考察 SMARW 方法与其他基于随机行走的抽样方法在可扩展性和无偏性方面的性能.具体考察的典型随机行走方法包括:

- Naïve 随机行走,每个节点转发消息到各个邻居概率相同.Naïve 方法是随机行走的最基础形式.
- Metropolis-Hastings 方法(下文缩写为 MH),在第 2.1 节进行了介绍,MH 引入自转移概率,提高了在不同拓扑结构下无偏抽样的能力.Stutzbach 等人^[14]直接利用 MH 方法,Awan 等人^[13]的方法是 MH 方法的变体.

模拟分为静态拓扑和动态网络波动环境下两部分.

(1) 由于非结构化 P2P 网络拓扑结构没有统一的标准,考察静态拓扑类型包括:

- Erdos-Renyi:最简单的随机图,以一定概率连接任意节点对.
- Watts-Strogatz:具有“小世界”特性的拓扑,有较高的聚集程度和低路径长度.
- Barabasi-Albert:也称为无尺度或幂率图,具有极端化的连接度分布.
- Gnutella 快照(snapshot):从 Stutzbach 等人^[10]获取的 Gnutella 拓扑快照中提取一个由超级节点组成的包括 156 493 个节点以及 1 768 315 条边的拓扑图.

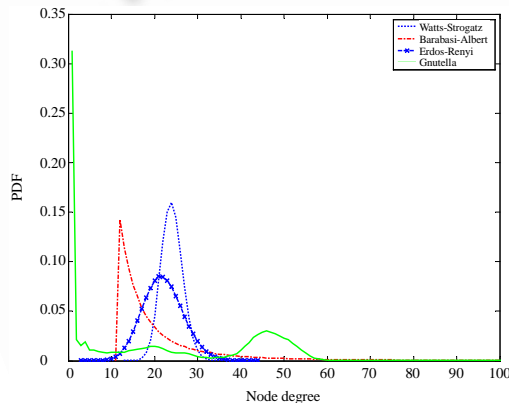


Fig.3 PDF distributions of topologies

图 3 不同拓扑连接度 PDF 分布

每次抽样的跳步数为 40(保证跳步数 $>\log n$).其中,SMARW 方法设定代理节点数目为 10.

A. DAM 过程收敛速度比较

考察 DAM 过程和 Awan 等人^[13]提出的 RWD 方法收敛速度.DAM 和 RWD 均通过分布式的降低自转移概率以提高随机行走的收敛速度.设定 Awan 等人^[13]推荐的 RWD 参数 $\delta=0.025, \rho=1000$,由于 RWD 方法在自转移概率 $<\delta$ 或没有相互更新的逻辑邻居节点时,即停止更新过程.故统计 RWD 方法自转移概率 $<\delta$ 的概率.结果如图 4 所示.

结果显示 DAM 过程能够迅速地降低自转移概率,平均轮数 <6 下 DAM 即达到稳定状态,而 RWD 方法在平均轮数 <15 时自转移概率变化较小;同时,图 4 显示 SMARW 方法在执行轮数 >6 (RWD 方法平均轮数 >45)后自转移概率 >0 节点比例即接近稳定状态,且小于对应 RWD 方法稳定时节点比例,显示 DAM 具有较高的收敛速度.同时,图 4 显示 Gnutella 快照中 DAM 和 RWD 达到稳定状态时自转移概率 >0 (或 $<\delta$)的节点比例分别约为 50%和 70%,这是由于 Gnutella 快照中存在超过 30%的节点连接度 <5 (有的甚至为 2),这些节点组成了大部分的自转移概率 >0 (或 $>\delta$)的节点.

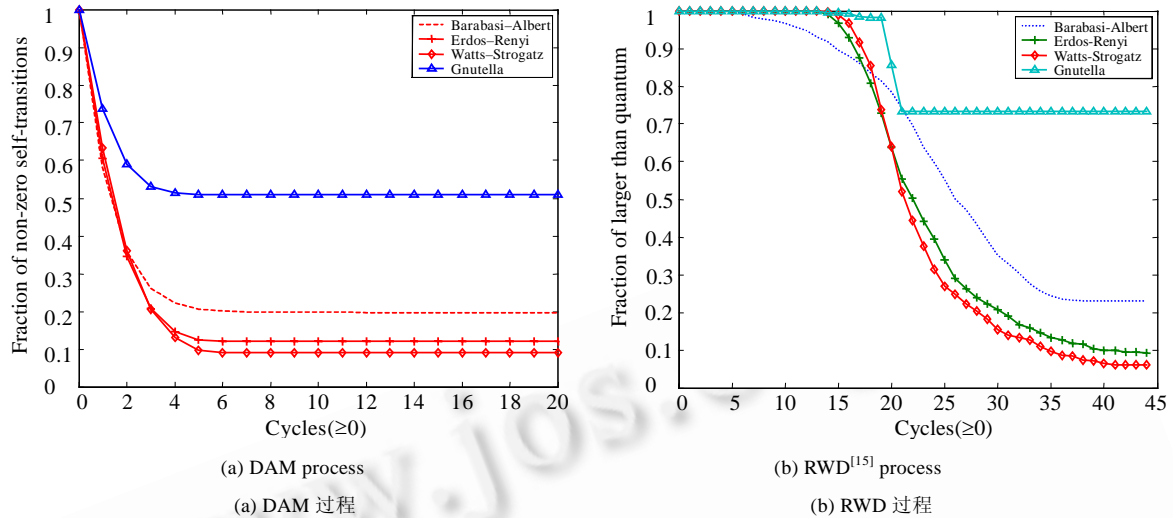


Fig.4 Convergence rate comparison between DAM and RWD process

图 4 DAM 与 RWD 过程收敛速度比较

B. 均匀随机度检验

利用 Kolmogorov-Smirnov 检验(KS 检验)对比相同跳步下 SMARW 方法与不同随机行走方法抽样的无偏程度.首先,离线的基于伪随机数发生器生成 $50 \times |V|$ 次的抽样作为理性情况下的抽样分布.为了对比 DAM 方法与 MH 方法在随机行走收敛速度方面的差别,分别考察基于 DAM 和 MH 的 MPRW 方法 KS 适配度差异.KS 检验比较实际频数与理论频数的累积概率间的差距,找出最大距离 D (其中 $D = \max\{\text{抽样概率分布} - \text{理想情况下概率分布}\}$),然后根据 D 值来判断实际频数分布是否服从理论频数分布.

KS 检验结果见表 1.在 5% 显著性水平下,KS 检验的临界值为 4.86×10^{-4} ,大于该临界值的抽样记录集合被认为缺乏足够的随机度.结果显示在各类拓扑结构下,MPRW 抽样记录适配度均低于临界值,说明 MPRW 具有较高的接近均匀随机分布的能力;而 Naïve 和 Metropolis-Hastings 方法缺乏稳定性,在某些拓扑结构下抽样记录被拒绝.同时,MPRW 方法在利用 DAM 方法下相对于基于 MH 方法具有相对较高的适配度,平均约为 1.5 倍,证明了 DAM 能够提高随机行走接近稳态分布的收敛速度.在下面的实验中,均采用 MPRW+DAM 的方式执行 SMARW 方法.

Table 1 KS test of different random walk methods under static topology environment

表 1 静态拓扑环境下不同行走方法的 KS 检验

	Watts-Strogatz	Erdos-Renyi	Barabasi-Albert	Gnutella
Naïve	2.3×10^{-3}	9.4×10^{-3}	2.2×10^{-3}	4.9×10^{-3}
MH	3.4×10^{-4}	5.8×10^{-5}	2.1×10^{-4}	5.6×10^{-4}
MPRW+MH	3.3×10^{-5}	5.2×10^{-5}	1.9×10^{-4}	4.7×10^{-4}
MPRW+DAM	2.6×10^{-5}	4.8×10^{-5}	3.6×10^{-5}	4.5×10^{-5}

C. 负载均衡

(C.1) 转发消息开销分布

统计每个节点转发随机行走消息的数目,给出节点转发消息开销的 CCDF 分布曲线,CCDF 定义为 $f(x)=1-P(X\leq x)$,CCDF 曲线能够较好地反映分布函数的尾部波动情况.如图 5 所示,结果显示在各类拓扑结构下 SMARW 方法消息开销分布均具有较高的“截断”特性,绝大部分集中在 2 000 左右,与第 3.3 节定理 5 分析的每个节点转发消息的期望值 $mL/N=(40\cdot 50|V|)/|V|=2000$ 非常接近,证明了 SMARW 能够完全消除转发随机行走消息的网络通信热点.而 Naïve 和 Metropolis-Hastings 方法的行走消息转发开销分布均具有较长的尾部延伸,说明这两种随机行走方法确实存在较大的行走消息转发的不均衡程度.

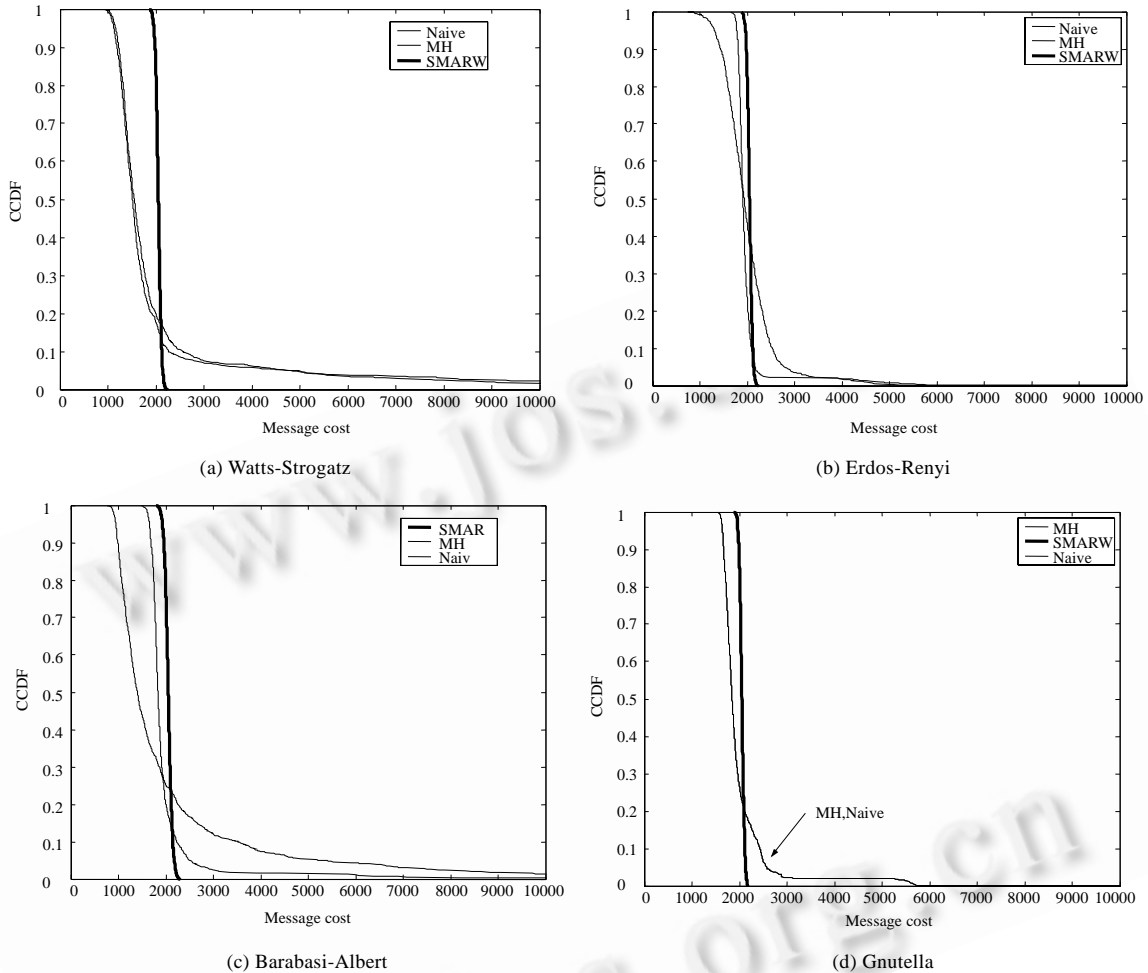


Fig.5 The CCDF distribution of message costs under different topologies

图 5 不同拓扑结构中节点转发消息开销的 CCDF 曲线

(C.2) 节点不均衡度 LIF 分布

统计在各类拓扑结构下各种行走方法的系统负载失衡度 LIF 值.回顾第 2.2 节 LIF 定义为转发行走消息开销最高值与转发行走消息开销分布中间值之比,显然最优的系统负载均衡情况下, $LIF=1$.如图 6 所示,结果显示在 3 类行走方法中 SMARW 方法具有最低的 LIF 值,约为 1,这充分证明了 SMARW 方法具有较高的负载均衡程度:每个节点的转发消息开销大致相同,接近最优的系统负载均衡情况.而 Naïve 和 Metropolis-Hastings 的 LIF 值均超过 25,显示这两种随机行走下的确存在严重的转发行走消息的网络通信热点.

(2) 考察网络波动环境下 3 种抽样方法(MH, Naïve, SMARW) 的差错程度及扩展性. 针对类似 Gnutella 网络中节点连接度进行抽样. 设定任意节点对之间延迟(RTT)均匀、随机地基于 King 方法^[17]来获取, 消息传输延迟为 $2 \times \text{RTT}$. 设定节点邻居数目在 [6, 100] 范围均匀地随机选择. 新节点加入服从 Poission 分布, 平均到达速率 $\lambda=1\text{s}$, 节点在线时间服从指数分布, 平均在线时间 10 000s, 设定节点在 10s 内未收到邻居节点更新消息即断开对应逻辑邻居连接. 运行模拟器一段启动时间达到稳定状态后进行抽样. 设定抽样数目 κ 为 500 000. Naïve 方法和 MH 方法从一个节点处同时获取 κ 个抽样; SMARW 方法设定代理节点为 10, 分布式地获取 κ 个抽样.

A. 抽样差错程度

统计上述 3 种抽样方法连接度分布和在线时间分布抽样差错程度. 基于 Stutzbach 等人^[14]的方式, 选择 SMARW 方法获取 50% 比例抽样节点时刻的拓扑结构快照(snapshot)为参考拓扑, 以快照中所有节点的连接度分布和在线时间分布为对应参考分布. 结果如图 7 所示.

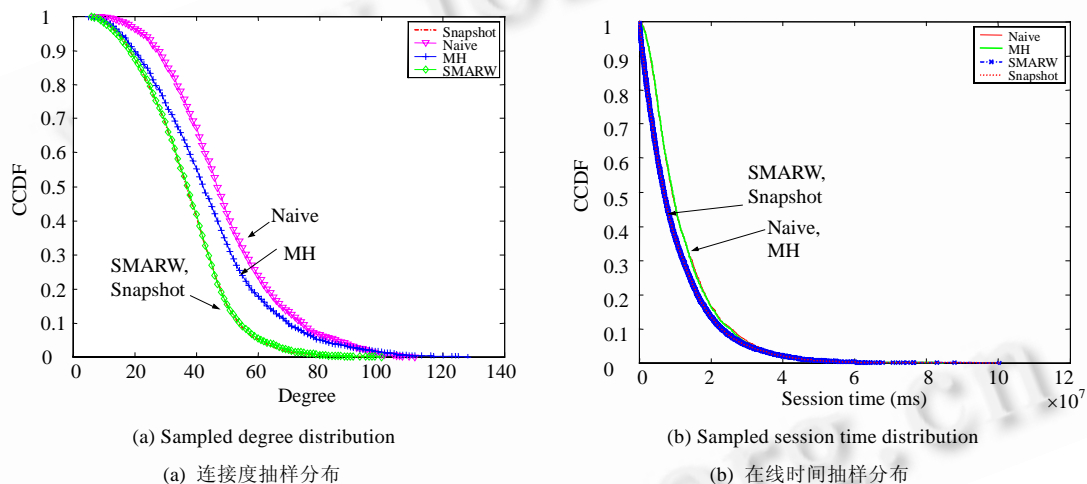


Fig.7 The CCDF distributions of degrees and session time of different sampling methods

图 7 不同抽样方法的连接度和在线时间的 CCDF 曲线

结果显示, Naïve 方法和 MH 方法获取的连接度和在线时间 CCDF 分布(此处及下文 CCDF 分布与上文 C.1 定义相同)均与快照中的实际 CCDF 分布存在偏差, 而 SMARW 方法与快照中实际分布基本吻合, 这是由于 SMARW 方法采用 DAM 过程进一步降低了节点自转移概率, 提高了随机行走的收敛速度, 从而减小了偏向于特定节点(连接度, 在线时间)的程度; 其次, 结果显示 Naïve 方法和 MH 方法获取的抽样 CCDF 分布曲线均高于快照对应的 CCDF 曲线, 说明上述两种方法偏向于连接度较高的节点及在线时间较高的节点.

B. 网络波动下负载均衡

统计不同抽样方法下所有曾经在线节点转发随机行走消息的数目. 结果如图 8 所示. 图 8(a) 显示了 3 类随机行走方法下节点转发随机选择消息的 CCDF 分布, 由于 Naïve 和 MH 方法从一个节点处同时发出多个抽样, 因此系统中节点转发消息负载较重, 平均 >5000 , 同时, 存在较长的尾部延伸, 即较大的负载不均衡程度; 而

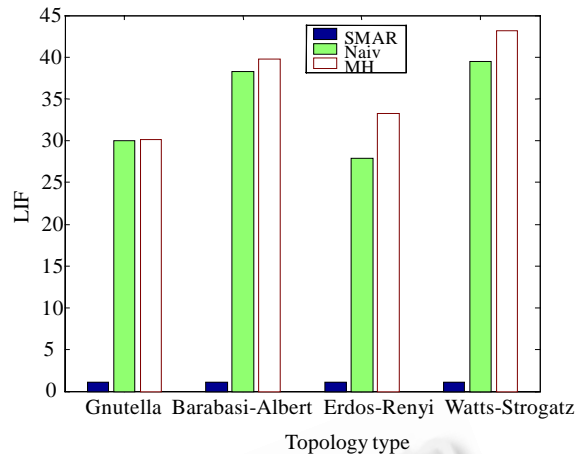


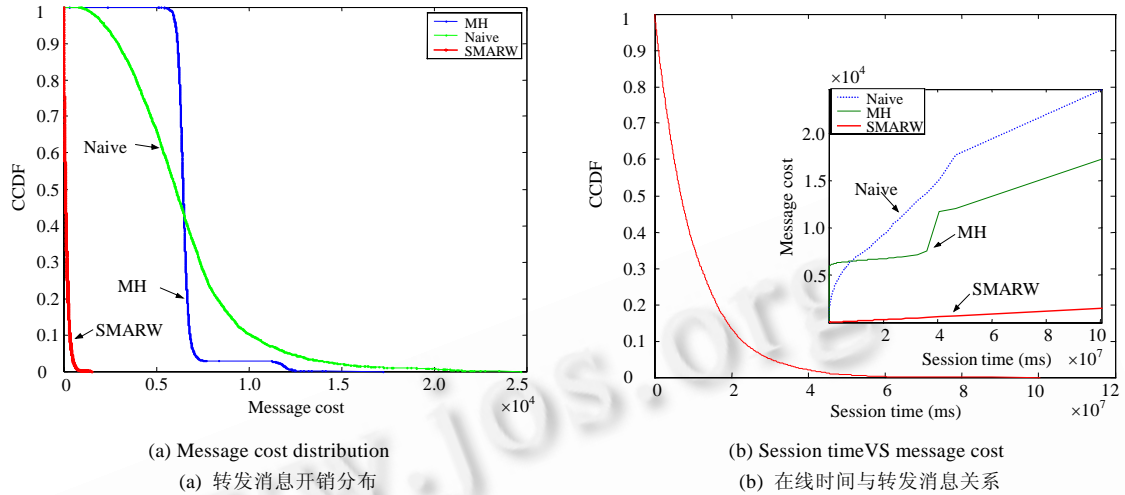
Fig.6 The LIF distribution of different random walk methods under different topologies

图 6 不同拓扑结构中 3 类行走方法 (MH, Naïve, SMARW) 的 LIF 值分布情况

SMARW 方法下节点转发消息负载极低,平均<1000,且具有较高的“截断”特性,显示 SMARW 方法具有较高的负载均衡度.图 8(b)显示节点在线时间与对应节点转发消息数目的关系曲线,结果显示,Naive 和 MH 方法随在线时间增大转发消息开销呈现波动现象,对应的关系曲线为非线性函数(称 $y=a \cdot x+b$ 为线性函数, a, b 为常量, y, x 为变量);而 SMARW 方法随在线时间增大转发随机行走消息开销呈线性增长,对应的关系近似为

$$y_{messageCost} = p_1 \cdot x_{SessionTime} + p_2, p_1 = 1.4854e^{-5}, p_2 = 1.7284.$$

显示 SMARW 方法极大地降低了对于不同在线时间节点的偏向.



(a) Message cost distribution
(a) 转发消息开销分布

(b) Session time VS message cost
(b) 在线时间与转发消息关系

Fig.8 The CCDF distribution of message costs of different methods under churn

图 8 网络波动环境下不同方法的转发行走消息分布

4 结论

针对非结构化 P2P 网络中可扩展的快速抽样问题,提出了基于多个 peer 自适应随机行走进行可扩展的无偏抽样方法——SMARW.理论分析和模拟测试验证了 SMARW 方法具有高度的负载均衡能力以及较高的无偏抽样能力. SMARW 方法最主要的贡献是能够实现在非结构化 P2P 网络中可扩展的大规模无偏抽样.获得的抽样节点信息可以被任意 peer 节点使用,因此,SMARW 方法可以作为一个提供无偏抽样节点信息的公共基础架构.进一步的工作包括分析在真实的网络波动情况下,SMARW 抽样方法可能面临的偏差以及在结构化 P2P 网络中 SMARW 方法的有效性.

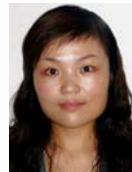
References:

- [1] Pandurangan G, Raghavan P, Upfal E. Building low-diameter peer-to-peer networks. IEEE Journal on Selected Areas in Communications, 2003,26:995-1002.
- [2] Vishnumurthy V, Francis P. On heterogeneous overlay construction and random node selection in unstructured P2P networks. In: Proc. of the INFOCOM. 2006. <http://www.cs.cornell.edu/~vivi/RandSelection.pdf>
- [3] Gkantsidis C, Mihail M, Saberi A. Random walks in peer-to-peer networks. In: Proc. of the INFOCOM. 2004. <http://www.cc.gatech.edu/~mihail/rwp2p04.pdf>
- [4] Verma S, Ooi WT. Controlling gossip infection pattern using adaptive fanout. In: Proc. of the ICDCS. 2005. 665-674.
- [5] Jelasity M, Guerraoui R, Kermarrec AM, Steen MV. The peer sampling service: Experimental evaluation of unstructured gossip-based implementations. In: Proc. of the 5th Int'l Middleware Conf. 2004. <http://www.inf.u-szeged.hu/~jelasity/cikke/middleware04.pdf>

- [6] Deshpande M, Xing B, Lazardis I, Hore B, Venkatasubramanian N, Mehrotra S. CREW: A gossip-based flash-dissemination system. In: Proc. of the ICDCS 2006. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1648832
- [7] Gummadi KP, Dunn RJ, Saroiu S, Gribble SD, Levy HM, Zahorjan J. Measurement, modeling, and analysis of a peer-to-peer file-sharing workload. In: Proc. of the Conf. on SOSP. 2003. http://www.cs.toronto.edu/~stefan/publications/sosp/2003/p2p_cache.pdf
- [8] Saroiu S, Gummadi PK, Gribble SD. Measuring and analyzing the characteristics of napster and gnutella hosts. Multimedia Systems Journal, 2003,9(2):170-184.
- [9] Pouwelse J, Garbacki P, Epema D, Sips H. The bittorrent P2P file-sharing system: Measurements and analysis. In: Proc. of the Int'l Workshop on Peer-to-Peer Systems (IPTPS). 2005.
- [10] Stutzbach D, Rejaie R, Sen S. Characterizing unstructured overlay topologies in modern P2P file-sharing systems. In: Proc. of the Internet Measurement Conf. 2005. 49-62.
- [11] Zhang H, Goel A, Govindan R. Improving lookup latency in distributed hash table systems using random sampling. IEEE/ACM Trans. on Networking, 2005,13(5):1121-1134.
- [12] Stutzbach D, Rejaie R. Improving lookup performance over a widely-deployed DHT. In: Proc. of the INFOCOM. 2006. <http://mirage.cs.uoregon.edu/pub/infocom06-kad.pdf>
- [13] Lovasz L. Random walks on graphs: A survey. Mathematical Studies, 1993,2:1-46.
- [14] Boyd S, Diaconis P, Xiao L. Fastest mixing Markov chain on a graph. SIAM Review, Problems and Techniques Section, 2004, 46(4):667-689.
- [15] Awan A, Ferreira RA, Jagannathan S, Grama A. Distributed uniform sampling in unstructured peer-to-peer networks. In: Proc. of the Hawaii Int'l Conf. on System Sciences. 2006. <http://ieeexplore.ieee.org/search/wrapper.jsp?arnumber=1579749>
- [16] Stutzbach D, Rejaie R, Duffield N, Sen S, Willinger W. On unbiased sampling for unstructured peer-to-peer networks. In: Proc. of the IMC. 2006.
- [17] Gummadi K, Saroiu S, Gribble S, King: Estimating latency between arbitrary Internet end hosts. In: Proc. of the SIGCOMM Internet Measurement Workshop (IMW 2002). 2002. 5-18.



符永铨(1983—),男,山东郓城人,硕士生,CCF 学生会员,主要研究领域为网络计算,P2P 网络.



周婧(1977—),女,博士生,主要研究领域为网络计算,虚拟现实.



王意洁(1971—),女,博士,教授,博士生导师,CCF 高级会员,主要研究领域为网络计算,数据库技术,移动计算.