Brief Announcement: Node Sampling Using Centrifugal Random Walks*

Andrés SEVILLA[1], Alberto MOZO[2], and Antonio FERNÁNDEZ ANTA[3]

[1] Dpto Informática Aplicada, U. Politécnica de Madrid, Madrid, Spain
[2] Dpto Arquitectura y Tecnología de Computadores, U. Politécnica de Madrid, Madrid, Spain
[3] Institute IMDEA Networks, Madrid, Spain
{asevilla,amozo}@eui.upm.es antonio.fernandez@imdea.org

Abstract. We propose distributed algorithms for sampling networks based on a new class of random walks that we call *Centrifugal Random Walks* (CRW). A CRW is a random walk that starts at a source and *always* moves *away* from it. We propose CRW algorithms for connected networks with arbitrary probability distributions, and for grids and networks with regular concentric connectivity with distance based distributions. All CRW sampling algorithms select a node with the exact probability distribution, do not need warm-up, and end in a number of hops bounded by the network diameter.

1 Introduction

Sampling the nodes of a network is the building block of epidemic information spreading [3], and can be used to construct small world network topologies [1]. A classical technique to implement distributed sampling is to use gossiping among network nodes [2]. A second popular distributed technique is the use of random walks [5]. Unfortunately, in these approaches, the desired probability distribution is reached when the stationary distribution of a Markov process is reached. The number of iterations (or hops of a random walk) required to reach this situation (the warm-up time) depends on the parameters of the network and the desired distribution, but it is not negligible.

We present efficient distributed algorithms to implement a sampling service. The basic technique used for sampling is a new class of random walks that we call *Centrifugal Random Walks* (CRW). A CRW starts at a network node, called the *source*, and *always* moves *away* from it. The sampling process in a *CRW algorithm* works essentially as follows. A CRW always starts at the source node. When the CRW reaches a node x (initially the source s), the CRW stops and selects that node with a given *stay probability*. If the CRW does not stop at x, it jumps to a neighbor of x that is farther away from the source than x. (The probability of jumping to each of these neighbors is not necessarily the same.)

Using this general approach, we firstly propose a CRW algorithm that samples *any* connected network with *any* probability distribution (given as nodes' weights). Before starting the sampling, a preprocessing phase is required. This preprocessing involves building a minimum distance spanning tree (MDST) in the network, and using this tree for efficiently aggregating the nodes' weights. Once the preprocessing is completed, any

* Work partially supported by Comunidad de Madrid (S2009TIC-1692) and Spanish MICINN (TEC2011-29688-C02-01). Thanks to Husain Ibraheem for useful suggestions.

node in the network can be the source of a sampling process, and multiple independent samplings with the exact desired distribution can be efficiently performed. Since the CRW used for sampling follow the MDST, they take at most D hops (where D is the network diameter).

Secondly, CRW algorithms without preprocessing are proposed when the probability distribution is distance-based (i.e., all the nodes at the same distance in hops from the source are selected with the same probability). The first distance-oriented CRW algorithm we propose samples with a distance-based distribution in a grid. In this network, the source node is at position $(0, 0)$ and the grid contains all the nodes that are at a distance no more than the radius R from the source. The algorithm we derive assigns a stay probability to each node that only depends on its distance from the source. However, the hop probabilities depend on the position (i, j) of a node and the position of the neighbor to which the CRW can jump. Since every jump of the CRW in the grid moves one hop away from the source, the sampling is completed after at most R hops.

For the general case of any connected network, we can picture nodes at each distance k from the source as positioned on a ring. The center of all the rings is the source, and the radius of each ring is one unit larger than the previous one. Using this graphical image, we refer the networks of this family as *concentric rings networks*. We have proposed a CRW algorithm that samples with distance-based distributions in concentric rings networks *with uniform connectivity*. These are networks in which all the nodes in each ring k have the same number of neighbors in ring $k - 1$ and the same number in ring $k + 1$. Like the grid algorithm, this one samples in at most R hops, where R is the number of rings. To deal with concentric rings networks with no uniform connectivity, we propose a distributed algorithm that, if it completes successfully, builds an overlay network that has uniform connectivity. In the resulting network, the algorithm for uniform connectivity can be used. We have found via simulations that this algorithm succeeds in building the desired overlay network in a significant number of cases.

In summary, CRW algorithms can be used to implement an efficient sampling service because, unlike previous Markovian, (e.g., random walks and gossiping) approaches, (1) they always finish in a number of hops bounded by the network diameter, (2) select a node with the *exact probability distribution*, and (3) do not need warm-up (stabilization) to converge to the desired distribution. Additionally, in the case that preprocessing is needed, this only has to be executed once, independently on the number of sources and the number of samples taken from the network. More details can be found at [4].

References

1. F. Bonnet, A.-M. Kermarrec, and M. Raynal. Small-world networks: From theoretical bounds to practical systems. In *OPODIS*, 2007.
2. M. Jelasity, S. Voulgaris, R. Guerraoui, A.-M. Kermarrec, and M. van Steen. Gossip-based peer sampling. *ACM Trans. Comput. Syst.*, 25(3), 2007.
3. D. Kempe, J. M. Kleinberg, and A. J. Demers. Spatial gossip and resource location protocols. *J. ACM*, 51(6):943–967, 2004.
4. A. Sevilla, A. Mozo, and A. Fernández Anta. Node sampling using drifting random walks. *CoRR*, abs/1107.1089v2, 2012.
5. M. Zhong and K. Shen. Random walk based node sampling in self-organizing networks. *SIGOPS Oper. Syst. Rev.*, 40:49–55, July 2006.