Non-Markovian Monte Carlo on Directed Graphs

Chul-Ho Lee

Florida Institute of Technology
cleee@fit.edu

Min Kang

North Carolina State University
mkang2@ncsu.edu

Do Young Eun

North Carolina State University
dyeun@ncsu.edu

ABSTRACT

Markov Chain Monte Carlo (MCMC) has been the de facto technique for sampling and inference of large graphs such as online social networks. At the heart of MCMC lies the ability to construct an ergodic Markov chain that attains any given stationary distribution π , often in the form of random walks or crawling agents on the graph. Most of the works around MCMC, however, presume that the graph is undirected or has reciprocal edges, and become inapplicable when the graph is directed and non-reciprocal. Here we develop a similar framework for directed graphs called Non-Markovian Monte Carlo (NMMC) by establishing a mapping to convert π into the quasi-stationary distribution of a carefully constructed transient Markov chain on an extended state space. As applications, we demonstrate how to achieve any given distribution π on a directed graph and estimate the eigenvector centrality using a set of non-Markovian, history-dependent random walks on the same graph in a distributed manner. We also provide numerical results on various real-world directed graphs to confirm our theoretical findings, and present several practical enhancements to make our NMMC method ready for practical use in most directed graphs. To the best of our knowledge, the proposed NMMC framework for directed graphs is the first of its kind, unlocking all the limitations set by the standard MCMC methods for undirected graphs.

ACM Reference Format:

Chul-Ho Lee, Min Kang, and Do Young Eun. 2019. Non-Markovian Monte Carlo on Directed Graphs. In *ACM SIGMETRICS / International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS '19 Abstracts)*, June 24–28, 2019, Phoenix, AZ, USA. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3309697.3331510>

1 BACKGROUND AND MOTIVATION

MCMC is a general technique for sampling from an arbitrary probability distribution by constructing a Markov chain that achieves the target distribution as its stationary distribution. By definition, every MCMC method requires the so-called balance condition, i.e., $\pi(j) = \sum_i \pi(i)P_{ij}$ for all j , where P_{ij} is the transition probability from state i to j and $\pi = [\pi(1), \pi(2), \dots, \pi(n)]$ is the stationary distribution. Most popular MCMC methods such as Metropolis-Hastings algorithm and Gibbs sampler, often referred to as Glauber dynamics, however, require a simple yet more restrictive condition,

called *detailed balance condition*, i.e., $\pi(i)P_{ij} = \pi(j)P_{ji}$ for all i, j . The detailed balance condition is generally preserved under a proposal and acceptance framework for state transitions. Assuming the current state i , a transition to state j is proposed with probability Q_{ij} , which is the transition probability of an irreducible (proposal) Markov chain, and then accepted with acceptance probability γ_{ij} . This two-step process ensures the convergence to the desired stationary distribution π via achieving the detailed balance condition, i.e., $\pi(i)Q_{ij}\gamma_{ij} = \pi(j)Q_{ji}\gamma_{ji}$ for all i, j . The detailed balance condition imposes a serious restriction, named ‘reciprocity’ condition, that $Q_{ij} > 0$ if and only if $Q_{ji} > 0$.

The detailed balance condition is a local condition that eases the construction of a Markov chain in MCMC methods, but is also criticized as a source of slow convergence to equilibrium and high sampling variability because every transition should be balanced out with its reverse transition. Thus, some recent advances in MCMC methods have been made to remove the requirement of the detailed balance condition in order to accelerate the convergence speed or improve sampling efficiency. A popular class of strategies is to transform a given reversible Markov chain to its related nonreversible chain by using the idea of lifting, nonbacktracking, or vorticity. Nonetheless, they all require the reciprocity condition, which is carried over from the original reversible chain before the transformation.

On the other hand, MCMC also fits into the problem of sampling from a target probability distribution over a large, *unknown* graph, or a huge state space whose exploration can only be done by traversing along the edges between nodes (states). The sampling problem with such limited access to the space often arises when sampling and estimating structural properties of various types of large networks, including peer-to-peer networks and online social networks, whose complete information is generally *unavailable* in advance. MCMC has then been a key enabler to obtain a sequence of samples by constructing an ergodic Markov chain or ‘random walk’ on a given graph, whose equilibrium distribution equals the target distribution, with only a limited and *local* knowledge of the graph structure.

Most of the MCMC methods for constructing random walks on graphs, however, become *inapplicable to directed* graphs, though many real-world networks can still only be represented by directed graphs. The directed graphs naturally violate the reciprocity condition, i.e., there is at least one directed edge (i, j) for which $Q_{ij} > 0$ but $Q_{ji} = 0$ due to the absence of its reverse directed edge (j, i) . The lack of reciprocity condition simply prohibits us from using the usual MCMC methods and even obtaining a closed-form expression of the stationary distribution of a given Markov chain on a directed graph. Thus, the stationary distribution has been merely estimated based on the visit counts obtained by running the Markov chain on a directed graph without knowing the explicit form of the distribution.

Do Young Eun was supported in part by National Science Foundation under Grant No. CNS-1423151 and No. CNS-1824518.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGMETRICS '19 Abstracts, June 24–28, 2019, Phoenix, AZ, USA

© 2019 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-6678-6/19/06.

<https://doi.org/10.1145/3309697.3331510>

Figure 1: Summary of the NMMC method: With any ergodic, nonreversible Markov chain Q defined over the node set N of a directed graph $\mathcal{G}$, we construct a transient Markov chain P on $N^+ = N \cup \{0\}$ such that its QSD becomes our desired distribution π and then leverage non-Markovian, history-dependent random walks to estimate the desired QSD π .

2 NMMC: QSD MAPPING AND ESTIMATION

In this paper, we present a general algorithmic framework called Non-Markovian Monte Carlo (NMMC) that constructs a set of non-Markovian random walks whose historical empirical distribution converges to any desired probability distribution π over a directed, non-reciprocal graph, using only local information on the graph structure. We here provide a brief overview of the NMMC method and refer to the full version of this paper in [1] for the details.

Consider a finite, aperiodic, directed graph $\mathcal{G} = (N, E)$. For a desired probability distribution π defined over N , the main idea of the NMMC method is to convert any given ergodic, nonreversible Markov chain (e.g., simple random walk) on N with its transition matrix $Q = [Q_{ij}]_{i,j \in N}$ and unknown stationary distribution μ (satisfying $\mu = \mu Q$) into a related suitable transient Markov chain on an augmented state space $N^+ := N \cup \{0\}$, where the node set N is a set of transient (non-absorbing) states and ‘0’ is the (artificially added) absorbing state. This mapping is made in a way that the resulting quasi-stationary distribution becomes the target distribution π .

Suppose that we have a discrete time transient Markov chain $\{X_t\}_{t \geq 0}$ on the state space $N^+ = N \cup \{0\}$ with its transition matrix $P = [P_{ij}]_{i,j \in N^+}$, where $P_{00} = 1$, i.e., once the chain hits 0, it remains there forever. We have $P_{i0} > 0$ for at least one $i \in N$ to ensure that the chain is transient, and assume that this Markov chain is irreducible and aperiodic on the non-absorbing states N . Then, there exists a unique invariant distribution $\nu = [\nu(1), \nu(2), \dots, \nu(n)]$ under the Markovian dynamics conditioned to survive, which is called the quasi-stationary distribution (QSD), satisfying

$$\nu(j) = \mathbb{P}\{X_t = j \mid X_0 \sim \nu, X_t \neq 0\}$$

for all $j \in N$. This can be rewritten with $t = 1$ as, for any $j \in N$,

$$\nu(j) = \frac{\mathbb{P}\{X_1 = j \mid X_0 \sim \nu\}}{\mathbb{P}\{X_1 \neq 0 \mid X_0 \sim \nu\}} = \frac{\sum_{i \in N} \nu(i) P_{ij}}{1 - \sum_{i \in N} \nu(i) P_{i0}},$$

which reads as $\sum_{i \in N} \nu(i) P_{ij} = \lambda \nu(j)$, with $\lambda = 1 - \sum_{i \in N} \nu(i) P_{i0}$. Letting $\tilde{P} = [\tilde{P}_{ij}]_{i,j \in N}$ be the restriction of the (stochastic) matrix P only on N , we have $\tilde{P}\nu = \lambda \nu$, where $\tilde{P}$ is a sub-stochastic matrix.

Adopting the proposal and acceptance framework as in usual MCMC methods, we first use the transition probability Q_{ij} as a proposal distribution from the current state (node) i to j ($i, j \in N$). This proposed move is then ‘accepted’ with probability γ_{ij} , which will be determined shortly. Unlike usual MCMC methods, however,

with probability $1 - \gamma_{ij}$, this move is ‘rejected’, in which case the chain gets absorbed to ‘0’ and stays there forever. We thus have $P_{ij} = Q_{ij}\gamma_{ij}$ for $i, j \in N$, and $P_{i0} = \sum_{j \in N} Q_{ij}(1 - \gamma_{ij})$ for each $i \in N$, with $P_{00} = 1$. Recall that $\tilde{P}_{ij} = P_{ij}$ for $i, j \in N$.

The next step is to choose the acceptance probability $\gamma_{ij} \in (0, 1]$ such that $\sum_{i \in N} \pi(i) \tilde{P}_{ij} = \sum_{i \in N} \pi(i) Q_{ij} \gamma_{ij} = \kappa \pi(j)$ is satisfied for all $j \in N$, i.e., $\pi \tilde{P} = \kappa \pi$ for some $\kappa > 0$. Then, by the Perron-Frobenius theorem, the distribution π must be the unique left leading eigenvector of $\tilde{P}$ and thus coincides with the unique QSD ν of the transient chain P , corresponding to the largest eigenvalue $\lambda = \kappa$. Specifically, for a given (proposed) ergodic, nonreversible chain Q , we define $S_j := \{k \in N \mid Q_{kj} > 0\}$ for each $j \in N$. For each S_j , consider a probability distribution $\{\alpha_{kj}\}_{k \in S_j}$ such that $\alpha_{kj} > 0$ for all $k \in S_j$ and $\sum_{k \in S_j} \alpha_{kj} = 1$. We then have the following.

THEOREM 2.1. For any choice of $\{\alpha_{kj}\}_{k \in S_j}$ on each S_j , $j \in N$, if we set $\gamma_{ij} \propto \frac{\pi(j)}{\pi(i)} \frac{\alpha_{ij}}{Q_{ij}}$ such that $\gamma_{ij} \leq 1$ for any pair (i, j) with $Q_{ij} > 0$, the target distribution π becomes the QSD ν of the transient chain P .

Lastly, there remain how to efficiently estimate or approximate the target QSD $\nu = \pi$ and how to obtain samples from the QSD in a distributed manner. To this end, one may consider long-run averages by generating many trajectories for $\{X_t\}$ for a long time, only to find out that most of them get absorbed or trapped at the absorbing site ‘0’ since $\mathbb{P}\{X_t \neq 0\}$ goes to zero exponentially fast, leaving very few trajectories to use. Thus, this brute-force approach simply would not work. Instead, we adopt history-based processes with reinforcement or history-dependent (thus non-Markovian) random walks to estimate the target QSD π , which are amenable to practical implementations. That is, under our NMMC method, a set of non-Markovian random-walk agents are launched to crawl the graph, and each of them independently moves over $\mathcal{G}$ according to the transition probability $P_{ij} = Q_{ij}\gamma_{ij}$, before it gets absorbed. Upon the absorption, it is instantaneously relocated (or redistributed) to one of the previously-visited nodes according to its historical empirical distribution, whose specific form is given in [1]. The overall operation of the NMMC method is illustrated in Figure 1.

REFERENCES

- [1] Chul-Ho Lee, Min Kang, and Do Young Eun. 2019. Non-Markovian Monte Carlo on Directed Graphs. *Proc. ACM Meas. Anal. Comput. Syst.* 3, 1, Article 15 (March 2019), 31 pages.