Analysis of Multiple Random Walkers for Service Discovery in Fog Computing Network Environments

Konstantinos Skiadopoulos, Konstantinos Giannakis, and
Konstantinos Oikonomou
Department of Informatics
Ionian University
Corfu, Greece

Email: {kskiad, kgiann, okon}@ionio.gr

Ioannis Stavrakakis
National and Kapodistrian University of Athens,
Greece - Universidad Carlos III de Madrid and
IMDEA Networks Institute, Spain
Email: ioannis@di.uoa.gr

Abstract—The multiple random walkers mechanism is revisited in this paper for service discovery purposes considering modern dynamic network environments like fog computing, where there is a need for certain services (e.g., virtual network functions) to be available to the end users. Network coverage is analytically investigated here considering multiple random walkers in geometric random graph topologies of various densities. The analytical results derived here are shown to be in accordance with other results in the literature (e.g., coverage under multiple random walkers for fully connected network topologies). Moreover, when there exist certain coverage and time constraints (i.e., the minimum fraction of network nodes to be covered within a certain time period), the required minimum number of random walkers satisfying these constraints is also analytically derived. Simulation results demonstrate the effectiveness of multiple random walkers for service discovery purposes and support the claims and expectations of the analysis.

Index Terms—Multiple random walkers, fog computing, coverage, service discovery, virtual network functions, cover time, geometric random graphs.

I. INTRODUCTION

Fog computing has seen a significant growth over the last few years as an intermediate network between the offered cloud services and the end user, particularly due to the increased number of users under the Internet of Things (IoT) and the proliferation of available services, e.g., [1], [2]. Fog computing is a distributed paradigm that acts as an intermediate layer between cloud data centers and other devices (e.g., mobile, sensors etc.). It offers computing, networking, and storage facilities so that cloud-based services can be extended closer to the end user [3].

In such network environments, the available resources within the fog network can be utilized and benefit the (closeby) end user. The network resources are often offered in the form of *virtual machines* that instantiate the particular services that are hosted either by the cloud *data centers* or *fog devices* [4], [5]. An example of such services is *virtual network functions* [6] where traditional network functions are distributedly executed on virtual machines located either in the cloud or the fog network [7], [8]. Eventually, this is a highly dynamic environment where end users move while virtual machines and the offered network functions may also change

their locations, thus making the problem of *service discovery* a challenging one. Another example is when big streams of data from IoT devices need to find their way to fog nodes acting as gateways to cloud data centers. Finally, situations where fog nodes are obligated to forward part or all of its workload to its neighbor nodes (for efficiency and improved quality), could also been considered as case studies [9].

In order to cope with this challenge, the work presented here revisits the deployment of multiple random walkers given their capability to visit fast certain network areas and (certainly) faster than a single random walker [10], [11]. An important characteristic of the mechanism of random walkers is its simple implementation and its capability to cover a large network area in a small time period, even though covering the entire network may take too long due to the randomness of the process.

The performance of multiple random walkers has been investigated for various topologies (e.g., power-law, random graphs, grids, etc., see also Section II). Nevertheless, these topologies are not representative for the considered environment where users exploit network resources of the particular service instantiations that are close in terms of distance. In the sequel, geometric random graph topologies, where nodes are connected when the euclidean distance among them is smaller than a certain connectivity radius, are considered (due to the characteristic of distance) [12]. Each employed random walker starts randomly visiting its neighbor nodes attempting to discover a certain service (e.g., a network function) in this particular unstructured and highly dynamic environment. The number of nodes visited, or covered by the multiple random walkers, as well as cover time, are analytically investigated as a function of the connectivity radius.

The main contribution of this paper is that the derived analytical results can be applied for the general case regarding the connectivity radius (assuming connected networks) and not for specific instances, e.g., fully connected topologies [11]. For example, when the requirement is to cover a certain proportion of a specific network within a given time period, the analysis in this paper provides for the minimum number of random walkers that need to be employed in order to satisfy these constraints. The agreement among the analytical results of this

paper and previous works in the literature is also investigated here, showing that the current analysis effectively captures and extends the previously proposed analytical models. Extensive simulation results take place considering instances of large networks, confirming the claims of the analysis.

In the sequel, past related works are reported in Section II, while Section III introduces various definitions to describe the under study problem. The proposed model is analyzed in Section IV. The performance evaluation using simulation results is presented in Section V and, finally, the conclusions are drawn in Section VI.

II. PAST RELATED WORK

As early as 2012, Bonomi et al. [1] introduced the idea of fog computing, mentioning also certain scalability issues [13]. The case of mobile fog devices is studied by Hong et al. [4], while the concept of nano data centers is investigated by Valancius et al. [5]. Further information is included in corresponding surveys, e.g., [2], while a comprehensive survey about network functions virtualization can be found in [6].

There are many works about performance analysis of multiple random walkers. In the fundamental work of [14], Lovász presents an in-depth analysis of the properties of random walkers. Alon et al. in [10] present a thorough study regarding cover time, concluding that the use of many random walkers in parallel yields a speed-up in the cover time that is linear in the number of employed walkers. It is also shown that an exponential speed-up is possible, even though a logarithmic speed-up is experienced most of the times.

Avin and Ercal in [15] study covering issues for random geometric graphs, showing that multiple random walkers are faster than one. Cooper and Frieze study cover time on sparse [16] and regular [17] random graphs, whereas Patel et al. investigate hitting time for the case of multiple random walkers [18]. A study regarding cover time of multiple random walkers on regular random graphs [19] reaches similar conclusions with this work that also studies multiple random walks on random geometric graphs.

Efremenko and Reingold in [20] calculate lower and upper bounds for cover time of multiple random walks, whereas they study three alternatives for the starting nodes of the random walks: the worst starting nodes (those maximizing cover time), the best starting nodes, and starting nodes selected from the stationary distribution. Cover time is also studied in [21] by Elsässer and Sauerwald, with tighter bounds on the provided speed-up over the use of a single random walker. Ivaskovic et al. in [22] derive tighter bounds on the speed-up for cover time in the case of d-dimensional grids.

In [23] a study by Beraldi is presented regarding random walks with long jumps on wireless ad hoc networks. Note that a random walk with long jumps shares similarities with multiple random walkers. Zheng et al. in [24] present an efficient search mechanism that employs random walkers with self-replication technologies in order to reduce the time delay. In addition, the case of multiple random walkers with

various replication mechanisms is analyzed on fully connected networks by Oikonomou et al. in [11].

Cecor-Hillel et al. in [25] study methods for checking particular graph properties. One of their proposed methods refers to the use of multiple random walkers. In particular, two such walkers are initiated at each node, therefore there are 2N random walkers operating concurrently, where N is the network size. Recently, Berenbrin et al. presented an alternative method for realizing the concept of multiple random walkers [26]. The underlying mechanism is related to randomized rumor spreading and shares a lot with epidemic spreading models. Cover time is also investigated under the proposed scheme, showing that for any connected regular graph the cover time is $O(N \log N)$ with high probability.

III. NETWORK AND COVERAGE DEFINITIONS

The analysis in the sequel considers a geometric random graph [12] model as the underlying network topology. There are N nodes, uniformly distributed on a plain area sized $[0,\ldots,1]\times[0,\ldots,1]$, that is a unit square. Each pair of nodes is connected if their euclidean distance is equal to or smaller than the connectivity radius r_c . Considering only connected topologies, m random walkers start at time step t=0 from an arbitrarily selected node to cover the network nodes. Let $C_m(t)$ (to be referred to hereafter as coverage) be a stochastic variable representing the $number\ of\ nodes$ that have been visited by at least one random walker after t time steps. Each random walker independently selects a random neighbor node in order to move at the next time step.

Assuming one random walker in the network (thus, $C_1(t)$ nodes are covered until time step t), there is a question regarding the relation between $C_1(t)$ and $C_m(t)$. Motivated by the seminal work of Alon et al. [10] that a linear speed-up is possible in most cases as long as m is not significantly large (i.e., of logarithmic order), it is assumed here that m random walkers increase m times the number of covered nodes on geometric random graphs, or,

$$C_m(t) = C_1(mt). (1)$$

Simulation results in Section V support the claims of Eq. (1).

IV. ANALYSIS

Let $P(r_c)$ denote the probability for a random walker to select a node that has not been visited during the last nine steps, based on the calculation of the common neighborhood of the currently visited node with the previously visited ones, as given in [27]. i.e., $P(r_c) = \left(1 - \frac{2V}{N\pi r_c^2 - 1}\right)$, where V is a constant ≈ 2.838 . Given that coverage $C_m(t)$ equals $C_m(t-1)$ plus the number of nodes that are covered by the m random walkers at time step t,

$$C_m(t) = C_m(t-1) + mP(r_c) \left(\frac{N - C_m(t-1)}{N}\right)$$

$$= C_m(t-1) + mP(r_c) \left(1 - \frac{C_m(t-1)}{N}\right)$$

$$= C_m(t-1) + mP(r_c) - \frac{mP(r_c)C_m(t-1)}{N}.$$

Eventually, the following recursive expression of $C_m(t)$ is derived,

$$C_m(t) = mP(r_c) + C_m(t-1)\left(1 - \frac{m}{N}P(r_c)\right).$$
 (2)

A. Coverage analysis

The next step is to transform this recursive expression to a more tractable form. The solution of the simple first order difference equation $y_{n+1}=ay_n+b$ for $a\neq 1$, i.e., $y_n=a^ny_0+b\left(\frac{a^n-1}{a-1}\right)$, is considered next. Thus, in the case of Eq. (2), where $a=1-\frac{mP(r_c)}{N}$ and $b=mP(r_c)$,

$$C_m(t) = \left(1 - \frac{m}{N}P(r_c)\right)^t + mP(r_c)\left(\frac{\left(1 - \frac{m}{N}P(r_c)\right)^t - 1}{1 - \frac{m}{N}P(r_c) - 1}\right)$$
$$= \left(1 - \frac{m}{N}P(r_c)\right)^t + N\left(1 - \left(1 - \frac{m}{N}P(r_c)\right)^t\right).$$

Finally, coverage of $C_m(t)$ is given by,

$$C_m(t) = N - (N - 1) \left(1 - \frac{m}{N} P(r_c)\right)^t.$$
 (3)

B. Model equivalence

It has been shown in the literature [11] that for the case of a fully connected network and a large number of nodes N,

$$C_m(t) = N\left(1 - e^{-\frac{mt}{N}}\right). \tag{4}$$

The next step is to show that coverage as given by the model introduced here, i.e., Eq. (3), reduces to the model presented in [11], i.e., Eq. (4), for fully connected graphs. In particular, when the connectivity radius r_c is so large that the network is fully connected, then the number of neighbor nodes $\approx N$. Assuming a large N, then $N \approx N-1$ and $P_{r_c} \approx 1$ [27]. Eventually, Eq. (3) is written as,

$$C_m(t) = N\left(1 - \left(1 - \frac{m}{N}\right)^t\right). \tag{5}$$

It is enough to show that for large values of N, $\left(1-\frac{m}{N}\right)^t=e^{-\frac{mt}{N}}$, or, $\ln\left(1-\frac{m}{N}\right)^t=\ln\left(e^{-\frac{mt}{N}}\right)$, or, $t\ln\left(1-\frac{m}{N}\right)=-\frac{mt}{N}$, or, $\ln\left(1-\frac{m}{N}\right)=-\frac{m}{N}$, or, $\ln\left(1-\frac{m}{N}\right)=-1$. For simplicity, let $x=\frac{m}{N}$. Obviously, as N takes large values, x converges to 0. Therefore, it suffices to show that $\lim_{x\to 0}\frac{\ln(1-x)}{x}=-1$. It is trivial to show the latter expression using L'Hôspital's rule, i.e., $\lim_{x\to 0}\frac{\ln(1-x)}{x}=\lim_{x\to 0}\frac{\ln(1-x)'}{x'}=\lim_{x\to 0}-\frac{1}{1-x}=-1$.

C. Minimum number of random walkers

From an implementation point of view, in the considered fog computing environment, it is of practical importance to have prior knowledge of the minimum number of random walkers m that need to start moving in the network, when certain coverage and time constraints are given. From Eq. (3), it is possible to calculate the minimum number of required random walkers in order to achieve a network *coverage fraction* k

within a specific amount of time T. Substitution of $C_m(T) = kT$ in Eq. (3), eventually yields,

$$m = N \frac{1 - (1 - k)^{\frac{1}{T}}}{P(r_c)},\tag{6}$$

assuming $N \approx N-1$. It is interesting to see that as k increases, the minimum number of required random walkers m increases. On the other hand, as T increases, m decreases and for $T \to +\infty$ (i.e., no constraints at all), then $m \to 0$.

V. SIMULATIONS

In this section, simulation results are presented, demonstrating the fact that the proposed analytic model of multiple random walkers captures the system's behavior. A program is developed in Python 3.6.3, using the SciPy and NumPy libraries. Randomness is generated by the random number generator of Scipy (i.e., the Mersenne Twister pseudo-random number generator) using different seeds for each run.

Geometric random graph [12] topologies are considered with 10^4 nodes. All topologies range from marginally connected (i.e., $r_c = 0.017$) to rather dense networks (i.e., $r_c = 0.100$). Marginally connected networks are those which consist of only one component, although a slight decrease of the value of the connectivity radius r_c would lead to more than one component. Eventually, 18 different network configurations are constructed with $r_c = [0.017, 0.020, 0.025, ..., 0.100]$. For each network, m = [1, 2, 3, 5, 10, 20] random walkers are employed and for each value of m the presented results correspond to average values of ten independent runs.

A. Single vs. multiple random walkers

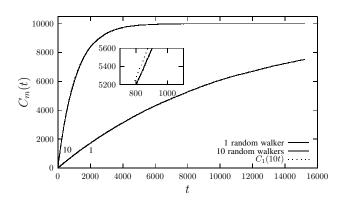


Fig. 1. Coverage for networks with $r_c=0.060,\,m=1,10$ random walkers. $C_1(10t)$ is also depicted.

Simulation results are used to demonstrate the effectiveness of the assumption behind Eq. (1). Fig. 1 depicts coverage for 1 and 10 random walkers in networks with connectivity radius $r_c = 0.060$ and the assumption that $C_{10}(t) = C_1(10t)$. It is obvious that the depicted simulation results are in accordance with Eq. (1) in the sense that ten random walker covers ten times the same number of nodes when compared to a single random walker for the same period of time.

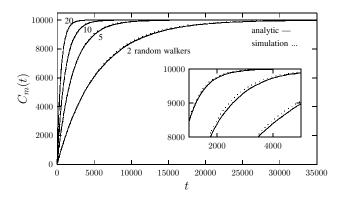


Fig. 2. Coverage for network topologies with $r_c = 0.060$.

B. Coverage

The next step is to demonstrate that Eq. (3) successfully captures the evolution of network coverage as a function of time. Fig. 2 depicts both analytic and simulation results regarding coverage for 2, 5, 10 and 20 random walkers employed in networks with $r_c=0.060$. As expected, the more the number of random walkers, the greater the coverage. In all cases the analytical results (dense lines) are in accordance with the results of the simulations (dotted lines).

C. Minimum number of random walkers

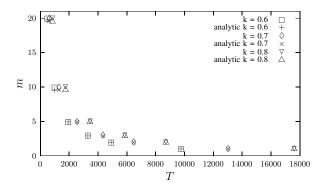


Fig. 3. Minimum number of required random walkers (m) for coverage fraction k=0.6,0.7,0.8 for network topologies with $r_c=0.060$ as a function of time period T.

The minimum number m of required random walkers to achieve coverage fraction k within T time steps is given by Equation (6). So, k, t, r_c, N are given as input to Eq. (6) and the suggested number of random walkers m is calculated for each case, given the time period T. In Fig. 3 the resulting minimum number of random walkers for both analysis and simulations, along with closeness is demonstrated for these cases. The largest deviation between analytic and simulation results is observed for marginally connected networks ($r_c = 0.017$) and it corresponds to one random walker.

D. Cover time

Cover time is defined as the number of steps required for m random walkers to cover the network. As already shown

in this work, the employment of multiple random walkers accelerates network coverage (see Eq. (1)). Fig. 4 depicts both analytic and simulation results as a function of the number of random walkers m for six different network topologies with r_c ranging from 0.017 (marginally connected) to 0.100 (dense network). The analytic results correspond to coverage for one random walker (i.e., $C_1(t)$) at time mt. Eventually, the comparison illustrated in Fig. 4 confirms again the validity of the assumption of Eq. (1), since for all depicted cases, it is observed that both simulation and analytical results are close. It is also noticed that as m increases, cover time decreases. Furthermore, as r_c increases, it is also evident that cover time decreases. This is expected, since as topologies become dense (i.e., r_c increases), a random walker is less likely to revisit the same network areas due to possible bottleneck links.

VI. CONCLUSIONS AND FUTURE WORK

Fog computing has emerged as a new distributed paradigm that acts as an intermediate layer between cloud data centers and user devices, offering various services. This new networking environment calls for new, efficient methods in order to leverage the capabilities it offers. In this direction, traditional approaches such as multiple random walkers were reconsidered and re-examined under the prism of fog computing, regarding coverage, service discovery (e.g., for virtual network functions) and information dissemination. In this work, multiple random walkers were employed in order to analytically investigate network coverage, while the equivalence to previous analytical models was also studied. The minimum number of random walkers required in order to satisfy certain coverage time constraints was also analytically derived. The results of the analysis were evaluated using simulations and it was shown that the analysis effectively captures the multiple random walkers behavior. It is left for future work to consider more realistic environments (e.g., traces from fog computing network topologies) in order to further demonstrate the applicability and effectiveness of the work presented here.

ACKNOWLEDGEMENTS

This work was supported in part by project "A Pilot Wireless Sensor Networks System for Synchronized Monitoring of Climate and Soil Parameters in Olive Groves," (MIS 5007309) which is partially funded by European andNational Greek Funds (ESPA) under the Regional Operational Programme "Ionian Islands 2014-2020." In addition, this work was supported in part by the European Commission as part of the ReCRED project (Horizon H2020 Framework Programme of the European Union under GA number 653417), by the Chair of Excellence UC3M - Santander Program and by the National and Kapodistrian University of Athens (S.A.R.G.).

REFERENCES

 F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the internet of things," in *Proceedings of the first edition of the* MCC workshop on Mobile cloud computing. ACM, 2012, pp. 13–16.

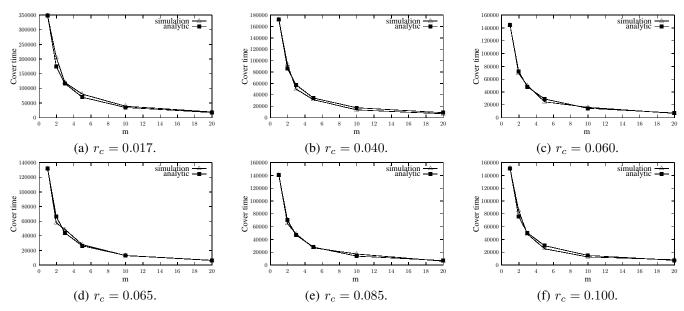


Fig. 4. Cover time as a function of the number m of random walkers for six different networks with r_c ranging from 0.017 (marginally connected network) to 0.100 (dense network).

- [2] S. Yi, C. Li, and Q. Li, "A survey of fog computing: concepts, applications and issues," in *Proceedings of the 2015 Workshop on Mobile Big Data*. ACM, 2015, pp. 37–42.
- [3] A. V. Dastjerdi, H. Gupta, R. N. Calheiros, S. K. Ghosh, and R. Buyya, "Fog computing: Principals, architectures, and applications," 2016.
- [4] K. Hong, D. Lillethun, U. Ramachandran, B. Ottenwälder, and B. Kold-ehofe, "Mobile fog: A programming model for large-scale applications on the internet of things," in *Proceedings of the second ACM SIGCOMM workshop on Mobile cloud computing*. ACM, 2013, pp. 15–20.
- [5] V. Valancius, N. Laoutaris, L. Massoulié, C. Diot, and P. Rodriguez, "Greening the internet with nano data centers," in *Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies*, ser. CoNEXT '09. ACM, 2009, pp. 37–48.
- [6] J. G. Herrera and J. F. Botero, "Resource allocation in NFV: A comprehensive survey," *IEEE Transactions on Network and Service Management*, vol. 13, no. 3, pp. 518–532, 2016.
- [7] R. Cohen, L. Lewin-Eytan, J. S. Naor, and D. Raz, "Near optimal placement of virtual network functions," in *Computer Communications* (INFOCOM), 2015 IEEE Conference on. IEEE, 2015, pp. 1346–1354.
- [8] E. Kavvadia, S. Sagiadinos, K. Oikonomou, G. Tsioutsiouliklis, and S. Aïssa, "Elastic virtual machine placement in cloud computing network environments," *Computer Networks*, vol. 93, no. Part 3, pp. 435–447, 2015, cloud Networking and Communications II.
- [9] Y. Xiao and M. Krunz, "QoE and power efficiency tradeoff for fog computing networks with fog node cooperation," in *INFOCOM 2017-IEEE Conference on Computer Communications, IEEE*. IEEE, 2017, pp. 1–9.
- [10] N. Alon, C. Avin, M. Koucky, G. Kozma, Z. Lotker, and M. R. Tuttle, "Many random walks are faster than one," in *Proceedings of the Twentieth Annual Symposium on Parallelism in Algorithms and Architectures*, ser. SPAA '08. ACM, 2008, pp. 119–128.
- [11] K. Oikonomou, D. Kogias, and I. Stavrakakis, "A study of information dissemination under multiple random walkers and replication mechanisms," ACM Proceedings of the Second International Workshop on Mobile Opportunistic Networking, p. 118, 2010.
- [12] M. Penrose, Random Geometric Graphs, ser. Oxford Studies in Probability. Oxford Univ. Press, 2004.
- [13] F. Bonomi, R. Milito, P. Natarajan, and J. Zhu, Fog Computing: A Platform for Internet of Things and Analytics. Cham: Springer International Publishing, 2014, pp. 169–186.
- [14] L. Lovász, "Random walks on graphs: A survey," Combinatorics, Paul Erdos is eighty, vol. 2, no. 1-46, p. 4, 1993.

- [15] C. Avin and G. Ercal, "On the cover time of random geometric graphs," in *Automata, Languages and Programming*, L. Caires, G. F. Italiano, L. Monteiro, C. Palamidessi, and M. Yung, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 677–689.
- [16] C. Cooper and A. Frieze, "The cover time of sparse random graphs," Random Structures & Algorithms, vol. 30, no. 1-2, pp. 1–16, 2007.
- [17] C. Cooper, A. Frieze, and T. Radzik, "Multiple random walks in random regular graphs," SIAM Journal on Discrete Mathematics, vol. 23, no. 4, pp. 1738–1761, 2009.
- [18] R. Patel, A. Carron, and F. Bullo, "The hitting time of multiple random walks," SIAM Journal on Matrix Analysis and Applications, vol. 37, no. 3, pp. 933–954, 2016.
- [19] C. Cooper, A. Frieze, and T. Radzik, "Multiple random walks and interacting particle systems," in *International Colloquium on Automata*, *Languages, and Programming*. Springer, 2009, pp. 399–410.
- [20] K. Efremenko and O. Reingold, "How well do random walks parallelize?" in Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques. Springer, 2009, pp. 476–489.
- [21] R. Elsässer and T. Sauerwald, "Tight bounds for the cover time of multiple random walks," *Theoretical Computer Science*, vol. 412, no. 24, pp. 2623–2641, 2011.
- [22] A. Ivaskovic, A. Kosowski, D. Pajak, and T. Sauerwald, "Multiple Random Walks on Paths and Grids," in 34th Symposium on Theoretical Aspects of Computer Science (STACS 2017), ser. Leibniz International Proceedings in Informatics (LIPIcs), vol. 66. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2017, pp. 44:1–44:14.
- [23] R. Beraldi, "Random walk with long jumps for wireless ad hoc networks," Ad Hoc Networks, vol. 7, no. 2, pp. 294–306, 2009.
- [24] Q. Zheng, X. Lu, P. Zhu, and W. Peng, "An efficient random walks based approach to reducing file locating delay in unstructured P2P network." IEEE GLOBECOM, 2005.
- [25] K. Censor-Hillel, E. Fischer, G. Schwartzman, and Y. Vasudev, "Fast distributed algorithms for testing graph properties," in *International* Symposium on Distributed Computing. Springer, 2016, pp. 43–56.
- [26] P. Berenbrin, G. Giakkoupis, and P. Kling, "Tight bounds for coalescingbranching random walks on regular graphs," in *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 2018, pp. 1715–1733.
- [27] K. Skiadopoulos, K. Giannakis, and K. Oikonomou, "Random walker coverage analysis for information dissemination in wireless sensor networks," *Technologies*, vol. 5, no. 2, p. 33, 2017.