

# The dynamics of knowledge acquisition via self-learning in complex networks

Thales S. Lima, Henrique F. de Arruda, Filipi N. Silva, Cesar H. Comin, Diego R. Amancio, and Luciano da F. Costa

Citation: [Chaos](#) **28**, 083106 (2018); doi: 10.1063/1.5027007

View online: <https://doi.org/10.1063/1.5027007>

View Table of Contents: <http://aip.scitation.org/toc/cha/28/8>

Published by the [American Institute of Physics](#)

---

## Articles you may be interested in

[Chaos in Kuramuro oscillator networks](#)

Chaos: An Interdisciplinary Journal of Nonlinear Science **28**, 071102 (2018); 10.1063/1.5041444

[Overcoming network resilience to synchronization through non-fast stochastic broadcasting](#)

Chaos: An Interdisciplinary Journal of Nonlinear Science **28**, 071104 (2018); 10.1063/1.5044420

[Quantifying entropy using recurrence matrix microstates](#)

Chaos: An Interdisciplinary Journal of Nonlinear Science **28**, 083108 (2018); 10.1063/1.5042026

[The variation of invariant graphs in forced systems](#)

Chaos: An Interdisciplinary Journal of Nonlinear Science **28**, 083101 (2018); 10.1063/1.5026551

[Engineering chimera patterns in networks using heterogeneous delays](#)

Chaos: An Interdisciplinary Journal of Nonlinear Science **28**, 071103 (2018); 10.1063/1.5042133

[Predicting tipping points of dynamical systems during a period-doubling route to chaos](#)

Chaos: An Interdisciplinary Journal of Nonlinear Science **28**, 073102 (2018); 10.1063/1.5038801

---



# The dynamics of knowledge acquisition via self-learning in complex networks

Thales S. Lima,<sup>1</sup> Henrique F. de Arruda,<sup>1</sup> Filipi N. Silva,<sup>2,3</sup> Cesar H. Comin,<sup>4</sup> Diego R. Amancio,<sup>1,3</sup> and Luciano da F. Costa<sup>2</sup>

<sup>1</sup>*Institute of Mathematics and Computer Science, University of São Paulo, São Carlos, São Paulo 13566-590, Brazil*

<sup>2</sup>*São Carlos Institute of Physics, University of São Paulo, São Carlos, São Paulo 13566-590, Brazil*

<sup>3</sup>*School of Informatics, Computing and Engineering, Indiana University, Bloomington, Indiana 47408, USA*

<sup>4</sup>*Department of Computer Science, Federal University of São Carlos, São Carlos, São Paulo 13565-905, Brazil*

(Received 26 February 2018; accepted 16 July 2018; published online 8 August 2018)

Studies regarding knowledge organization and acquisition are of great importance to understand areas related to science and technology. A common way to model the relationship between different concepts is through complex networks. In such representations, networks' nodes store knowledge and edges represent their relationships. Several studies that considered this type of structure and knowledge acquisition dynamics employed one or more agents to discover node concepts by walking on the network. In this study, we investigate a different type of dynamics adopting a single node as the “network brain.” Such a brain represents a range of real systems such as the information about the environment that is acquired by a person and is stored in the brain. To store the discovered information in a specific node, the agents walk on the network and return to the brain. We propose three different dynamics and test them on several network models and on a real system, which is formed by journal articles and their respective citations. The results revealed that, according to the adopted walking models, the efficiency of self-knowledge acquisition has only a weak dependency on topology and search strategy. *Published by AIP Publishing.* <https://doi.org/10.1063/1.5027007>

**Complex networks have been employed to model a myriad of real systems and processes, including the dynamics of knowledge acquisition and transmission. In this study, knowledge is represented as a set of entities (nodes) connected by links (edges). The process of acquisition is performed by agents walking on this structure. The discovery of each new piece of knowledge is transmitted to a brain, which is placed on a single node from where walkers start their walks. Our results revealed that the efficiency of knowledge acquisition hardly depends on the structure and walk.**

## I. INTRODUCTION

Most matter and energy exchanges in nature can be represented and understood in terms of information<sup>1</sup> flowing between distinct agents and/or subsystems. Perhaps as a consequence, several developments in network science have shown that a great number of natural systems can be represented and modeled in terms of intricate graphs, which have been called *complex networks*. Thus, interesting real-world systems such as opinion and epidemic spreading,<sup>2-4</sup> transport, communications and language,<sup>5-8</sup> as well as the nervous system<sup>9</sup> have been approached by network science, yielding several interesting results. One particularly interesting research line concerns how information is acquired by agents in a complex system.<sup>10</sup> Preliminary investigations along this line have represented knowledge as being stored into nodes that are logically interconnected, while one or more agents get acquainted with this knowledge while moving along the network.<sup>11</sup> Therefore, knowledge integration is

performed internally to each agent. This situation is illustrated in Fig. 1(a).

Many real-world situations can be represented and modeled by using this agent-based approach. The efficient routing of information in a complex system, such as the Internet, requires specific strategies for transmitting the information between nodes.<sup>12</sup> When there is limited knowledge about the global connectivity of the system, the routing problem becomes a heuristic search for the optimal paths for communication.<sup>13</sup> For instance, Lee *et al.*<sup>14</sup> proposed a degree-biased random walk search for optimizing the information exchanges on peer-to-peer systems. Kim *et al.*<sup>15</sup> showed that self-avoiding walks can provide nearly optimal exploration of networks. The search and information exchange dynamics on networks has also been studied from the perspective of opinion diffusion on social networks.<sup>16</sup> In this case, nodes represent individuals that need to form an informed opinion about a given subject, which is done by gathering the opinion of other nodes in the network. For example, McCullen *et al.*<sup>17</sup> analyzed the propagation of innovation, showing that the level of adoption of a new concept or product strongly depends on the topology of the network.

Here, we consider a distinct type of agent-based dynamics. We analyze the case where the knowledge is integrated not by one or more agents, but is otherwise stored in a specific and fixed part of the system, corresponding to some kind of “brain” or “hq” (headquarter), and the information is gathered by sampling through the networks by agents or sensing pathways. Real-world related systems include the mapping of communities with shared interests in the World-Wide-Web (WWW) or the Internet, as well as consciousness acquisition

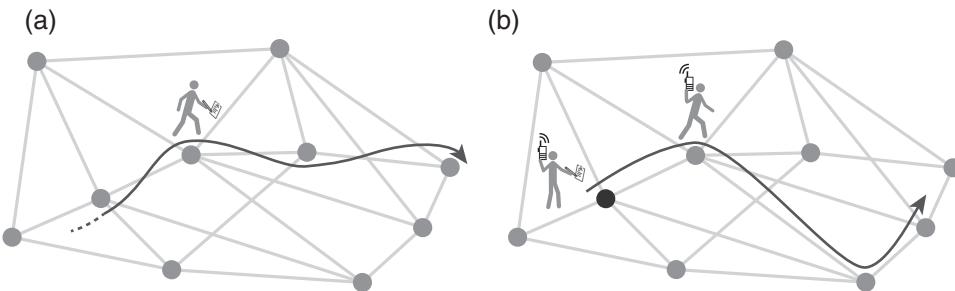


FIG. 1. Two possible cases of knowing a network. In the first case (a), an agent repeatedly departs from different nodes and performs random walks aimed at collecting knowledge about the network topology. In the second case (b), the collection of knowledge about the network is centered at a fixed node, here called *network brain*. An agent repeatedly departs from the *same* node, collecting information about the network topology. The latter situation is particularly interesting in the sense that the choice of the *brain* can influence strongly the knowledge acquisition efficiency.

by individuals with respect to the environment. In these situations, the information to be integrated can be considered as being part of the own organism, so that acquired knowledge becomes *self-knowledge*. For simplicity's sake, the central node where information is integrated in these systems is henceforth called the *network brain*.

Though more general knowledge acquisition situations, such as by multiple agents, are more often addressed in the research literature, the particular dynamics of self-knowledge represents an especially important problem in several real-world situations involving centralized monitoring and control. For instance, the self knowledge of an organism is required for its operation and survival. More specifically, muscles and organs of an individual are monitored and controlled continuously through the nervous system, having the brain as the major control and decision portion. Similarly, energy, communications, and transportation systems require the constant monitoring and control of their parts, which nowadays are mostly implemented through distributed computing systems, message exchanges, and autonomous agents that communicate with specific control unities.

Other analogous cases include traffic and weather monitoring, as well as social insects. All these systems require continuous self-awareness because they are subjected to changes which may require important response. Though these systems do not necessarily involve moving agents reporting to a central headquarter as in this work, their operation bears direct analogy in the sense that their respective communication mechanisms can often be well-approximated in terms of flows of information between agents at different positions and one or more central control unities. In addition to maintaining routine maintenance activity, these systems need to have mechanisms to detect and map topological changes in their own structure. For instance, the WWW needs to be continuously explored by worms in order to detect new pages and documents and, therefore, to provide a reasonably accurate map of its interconnections.

The type of centralized knowledge acquisition considered in this work, illustrated in Fig. 1(b), has seldom been addressed from the perspective of computational modeling. Yet, this type of systems entails some specific problems that are particularly interesting. One of the main such problems regards how influential is the topology around the brain on the knowledge acquisition speed. In case the topology is influential, this would imply special care while electing a

specific node to operate as brain or hq of the system whenever possible. The present work addresses precisely this type of centralized systems as well as the influence of topology on brain placement. We resort to complex network models and concepts in order to investigate how self-knowledge progresses in different types of networks and how this dynamics is influenced by the choice of node to act as the network brain.

In the investigations reported in this work, we consider the domain to be explored to be represented as a complex network. The knowledge exploration is performed by agents that move along the network through random walks and periodically report the information gathered to the network brain, where it is integrated and stored. These agents can be thought as independent and simultaneous, thus the knowledge of different walks is not transmittable between them. A number of different topologies are considered in order to allow us to infer to which extent the structure of the domain influences the self-exploration efficiency. At least as important as the topology in defining such an efficiency is the choice of node to act as brain and the type of probing dynamics.

Recent studies suggest that the topology of the network, and also around the starting node, tends to have little influence on the knowledge acquisition speed, at least for the considered models.<sup>11</sup> However, these results refer to moving agents as in Fig. 1. Because the random walks implemented in such cases tend to “forget” the initial node, the topology around these nodes is intrinsically less influential on knowledge acquisition. In the current work, however, the node where the knowledge is integrated is fixed, and all the several implemented random walks depart from this same node. Therefore, it could be expected that the topology surrounding these nodes plays a more decisive role on the knowledge acquisition speed.

This article is organized as follows. Section II describes some works on random walks. Section III presents the employed network and dynamics. In Sec. IV, we analyze the adopted dynamics and compare their efficiency according to different used parameters. Finally, in Sec. V we conclude the study and make some remarks regarding the results.

## II. RANDOM WALKS

One of the first studies regarding random walks was proposed by Pearson in 1905.<sup>18</sup> He aimed at modeling the movements of mosquito swarms (called walkers) in forests.

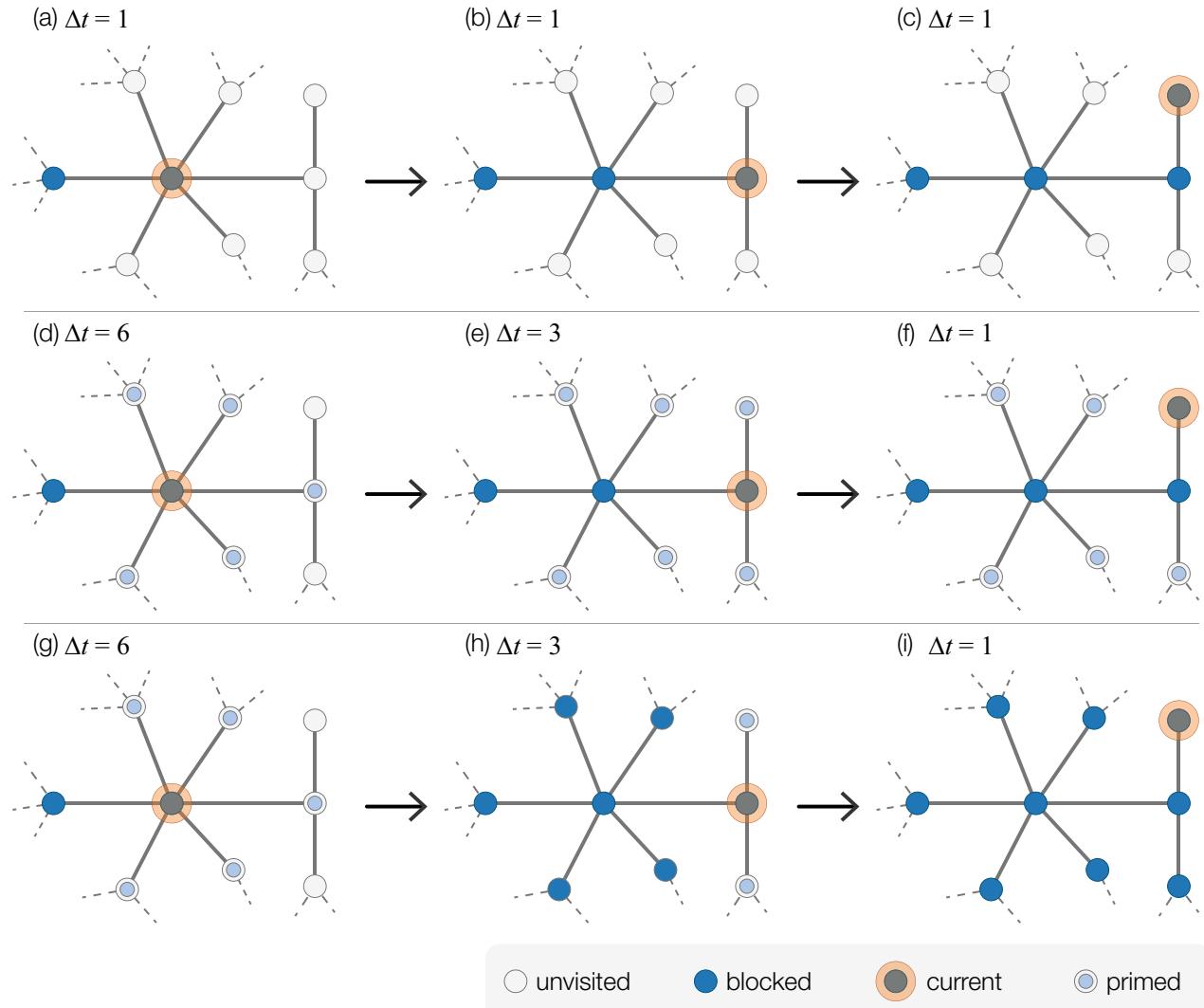


FIG. 2. Description of the changes of states in the three adopted walking models. The four states that a node can assume are shown in the legend, and include: *unvisited*, *primed* (a node that is known but can still be visited), *blocked* (a node that is known and cannot be visited), and *current* (the node where the agent is). (a)–(c) represent the self avoiding dynamics, where the only known nodes are those that have been visited already; (d)–(f) the extended self avoiding where all nodes that neighbor the path are also known but not avoided; and (g)–(i) the look-ahead self avoiding where all nodes that neighbor the path are known and are avoided.

He defined a dynamics in which for each iteration, the walker turns a random angle and walks in a straight direction at a fixed distance. This process is repeated many times. By considering this dynamics, Pearson proposed the problem of calculating the probability distribution of the distance traveled by the walker from the starting point. The answer was given by Lord Rayleigh.<sup>19</sup> Since then, many applications have benefited from the use of random walk models.<sup>20–25</sup> More recently, many studies took into consideration this dynamics simulated on complex networks,<sup>22</sup> in which the network topology is employed to restrict the walker movements. For instance, this dynamics has been used to model users navigating on the World-Wide-Web.<sup>26</sup> In this case, nodes represent web pages and edges indicate a hyperlink between web pages. Using the model, the web pages can be ranked according to the number of times they are visited by the walker. Such a ranking was called the PageRank algorithm.<sup>27</sup>

There are many different types of random walk dynamics on networks, the most basic, and the traditional being the

uniform random walk. In this case, the walker selects one of the neighbors of the current node with equal probability. Therefore, the transition probability from a node  $i$  to another node  $j$  only depends on the degree of node  $i$ . Given enough time, the walker visits every node of the network. The average number of steps taken by the walker to visit the entire network is called the cover time.<sup>28</sup> The uniform random walk tends to have relatively large cover times when compared to other random walk models. This is so because the walker tends to revisit many times the nodes that have already been visited. A modification in the dynamics, which can greatly reduce cover times, is to forbid the walker to revisit nodes that have already been visited, defining a *self-avoiding* walk.<sup>11</sup>

### III. MATERIALS AND METHODS

This section presents the adopted networks, which include three real-world cases and many theoretical models,

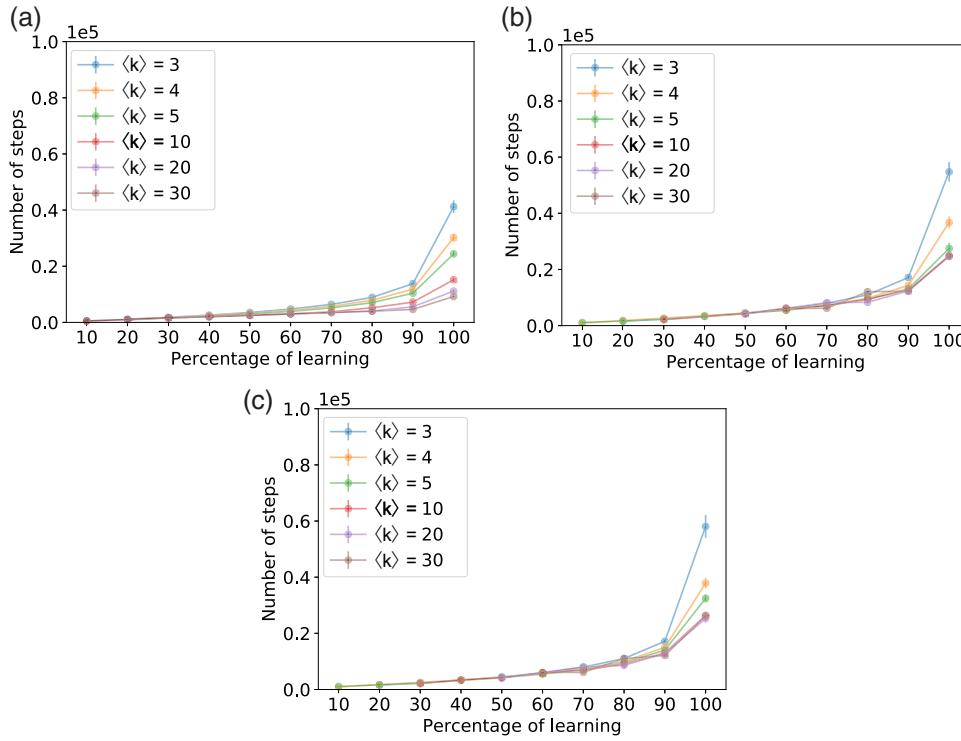


FIG. 3. Learning efficiency for the ER networks considering the three adopted dynamics. Each curve corresponds to a realization for a different average node degree  $\langle k \rangle$ , according to the legend. (a) Standard, (b) Extended, and (c) Look-ahead.

the topological measurements used to characterize their topology, and the implemented self-learning dynamics.

### A. Complex networks models

Networks with the same size and number of edges were generated for each different model studied. The models that were studied in this work were:

*Random Networks (Erdős-Rényi—ER):* This has been used as the reference model in complex networks research.

Each of its nodes has a probability  $p$  of connecting to each other node in the network.<sup>29</sup>

*Scale Free Networks (Barabási-Albert—BA):* This type of network is obtained by incrementally adding nodes with probability of connection proportional to the degree of the nodes already existing in the network. The most important property of this network model is the power law degree distribution and the hence implied presence of hubs.<sup>30</sup>

*Configuration Model (CM):* In this model, a random graph is created to fit a desired degree distribution.<sup>31</sup> The

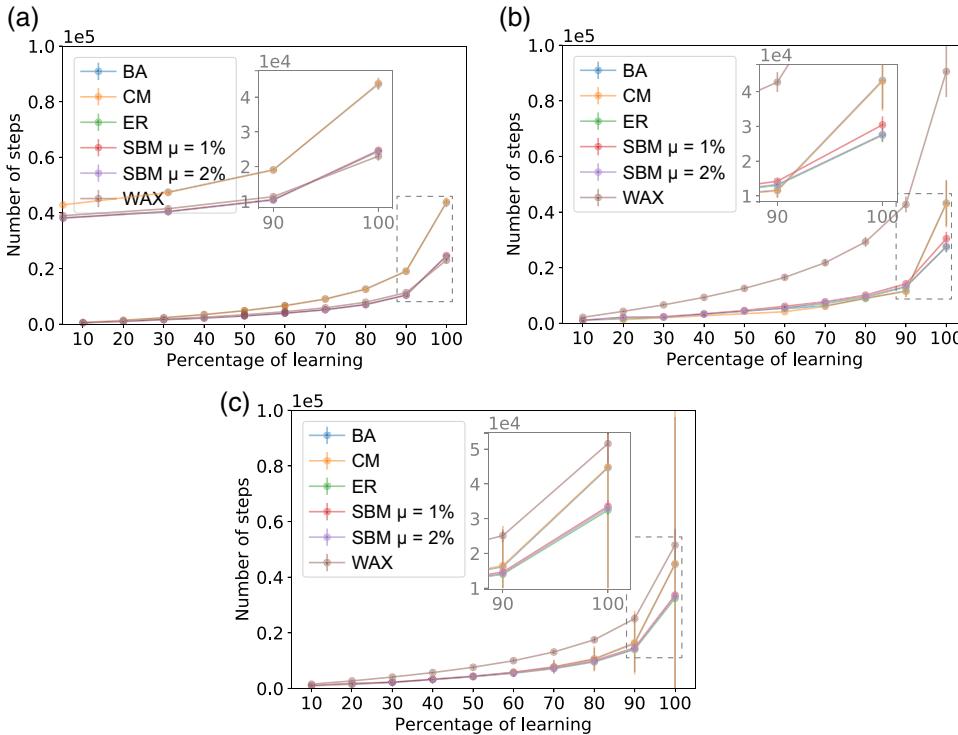


FIG. 4. Efficiency curves for the considered walk dynamics and network models. In all cases,  $\langle k \rangle = 5$ . (a) Standard, (b) Extended, and (c) Look-ahead.

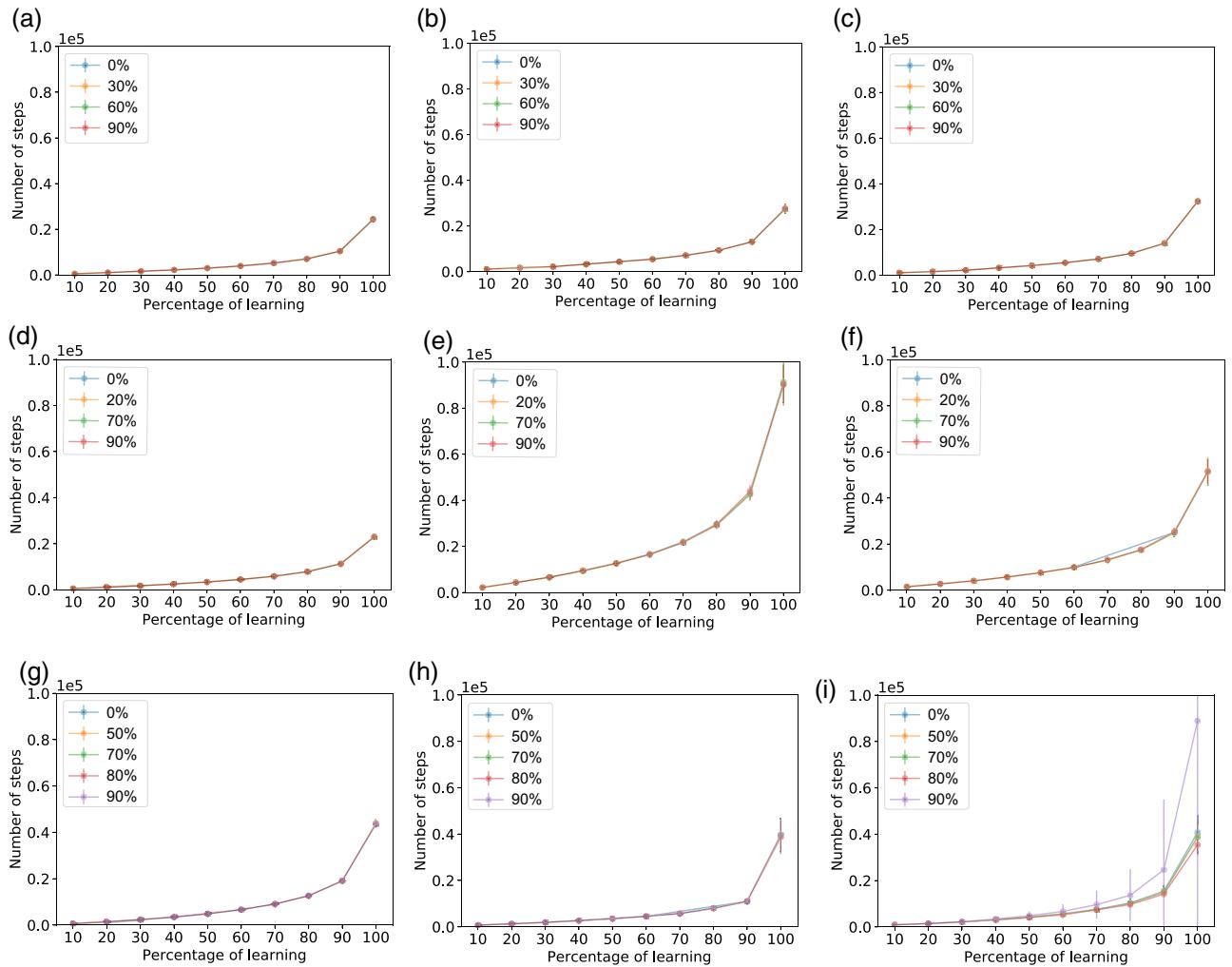


FIG. 5. Learning efficiency obtained for all the considered combinations of walk dynamics and network models. Each curve corresponds to a selection of the starting node having betweenness centrality higher than a certain percentile considering the entire network, as indicated in the legends. (a) ER Standard, (b) ER Extended, (c) ER Look-ahead, (d) WS Standard, (e) WS Extended, (f) WS Look-ahead, (g) BA Standard, (h) BA Extended, and (i) BA Look-ahead.

configuration model was used as a means to better understand some complex effects detected in the BA networks. The degree distributions used in this work were generated to follow a BA network.

*Small World Networks (Watts-Strogatz—WS):* This model is generated by rewiring connections in a regular toroidal lattice. It presents the combination of high clustering coefficient and low average minimal distance.<sup>32</sup> The standard rewiring probability used was 3%.

*Geographical Networks (Waxmann—WAX):* These are based on distributing points in the Cartesian plane. The nodes are evenly distributed, and each node can connect to each other with probability decreasing exponentially with the distance. For the Waxmann networks, we used a square space with random distribution of nodes in it.<sup>33</sup>

*Stochastic Block Model (SBM):* This model uses blocks of nodes and probabilities of connection between nodes from one block to another (even with itself).<sup>34,35</sup> Two models of SBM networks were adopted in this work, both with 10 blocks. Each block has the same probability  $\mu$  of connecting to the other blocks, the first model has a  $\mu = 1\%$  chance and the second  $\mu = 2\%$  chance.

All parameters referring to node connectivity were chosen to approximately reflect the desired average degree of the network, i.e., the probability of connection between two nodes in the ER network.

## B. Real-world networks

In order to investigate the behavior of the proposed dynamics on real systems, we adopted three distinct real-world networks: a citation network obtained from the Web of Science (WOS),<sup>36</sup> the California WWW (CAL) network,<sup>37,38</sup> and the US Power Grid (PG) network.<sup>32</sup>

The WOS network comprises articles returned by searching for complex network topics in journals indexed by WOS. Each node corresponds to an article, and a connection exists between a pair if they cite each other. Note that, in this case, we disregard the direction of the connections since we are interested in the relatedness between topics and not in citation dynamics. The network contains 11 063 articles and has average degree  $\langle k \rangle = 17$ . More information about this network can be found in Ref. 39.

The CAL network is a Web sub-graph obtained in 2002 by expanding the connections from a predefined set of

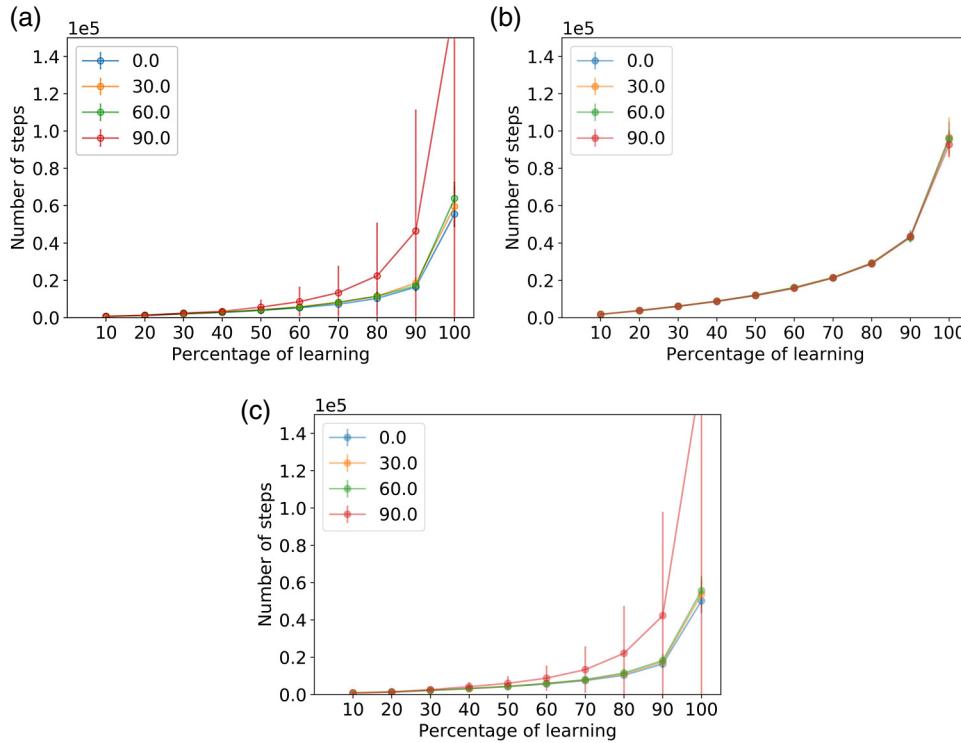


FIG. 6. Learning efficiency obtained for all the considered combinations of walk dynamics and network models. Each curve corresponds to a selection of starting node having centrality (accessibility or eigenvector centrality) higher than a certain percentile considering the entire network, as indicated in the legends. (a) BA Look-ahead (accessibility), (b) WS extended (accessibility), and (c) BA look-ahead (eigenvector).

domains using the query “California.” Each node corresponds to a website and connections represent the hyperlinks among them. It contains 9664 nodes and the average node degree is  $\langle k \rangle = 3.3$ .

Different from the other models, the PG network does not represent an information network. However, its structure resembles a geographic communication network. The PG network comprises 4941 nodes associated with power stations and generators. Each edge indicates a power line between nodes, leading to an average degree  $\langle k \rangle = 2.7$ .

### C. Walk models

With these networks in hand, we aim at understanding how long does it take for a random walker to fully discover the network. The efficiency is measured with the total number of steps taken to explore the network. Note that this efficiency measurement is strongly related to the concept of cover time for traditional random walks.<sup>28</sup> Three different walking models were used:

*Standard self-avoiding random walks:* for every move in these walks, the walker chooses the destination node randomly in the current neighborhood. The sole restriction on these walks is that the same node cannot be visited twice. A node is considered to be known if the walker has visited it at least once. Therefore, the number of steps of a walk using this dynamic is the number of nodes of the walk. Thus, the number of steps of each iteration, denoted by  $\Delta t$ , is always 1, as illustrated in Figs. 2(a)–2(c).

*Extended self-avoiding random walks:* in this model the walker behaves exactly like in the previous model, but the known region is considered to include not only the nodes that were covered along the walk, but also the neighbors of these nodes. This means that the number of steps in a walk

correspond to the sum of the degrees of each of its nodes, as seen in Figs. 2(d)–2(f). Note that here we consider that checking whether a node is already known by the walker is analogous to taking a step in the walk. In other words, the traveler spends effort/time when checking the status of neighbors. This is the reason why we denote the total number of steps as  $\Delta t$ .

*Look-ahead self-avoiding walks:* like in the previous model, here the walker will be considered to know the neighborhood of every visited node after departing from it. However, the agents are prohibited to proceed to any already known node. The number of steps of a walk is calculated in the same fashion as in the extended walk dynamics. This dynamics is illustrated in Figs. 2(g)–2(i).

Note that, for every model, the walks may eventually hit a dead end and will have to stop, prompting another walk to be started from the same starting point. This is repeated until all nodes have been discovered. For every new iteration the same starting node is used and the network is considered to be completely unvisited by the agent (the knowledge about the previous walks is always kept in the network brain). We decided to select only self-avoiding random walks in our analysis as such walks tend to optimize the exploration efficiency. This happens because the walker tends to visit an unvisited node at each new step, thus avoiding those previously visited nodes. The restart mechanism in the adopted walks also favors efficiency because it avoids traps.

Since every network model was created with the same number of nodes, but the average degree varies from 3 to 30, the probability of dead ends will vary considerably. To better grasp the influence of very long walks, an additional configuration was adopted in which the maximum number of steps before resetting the walk was set to 100.

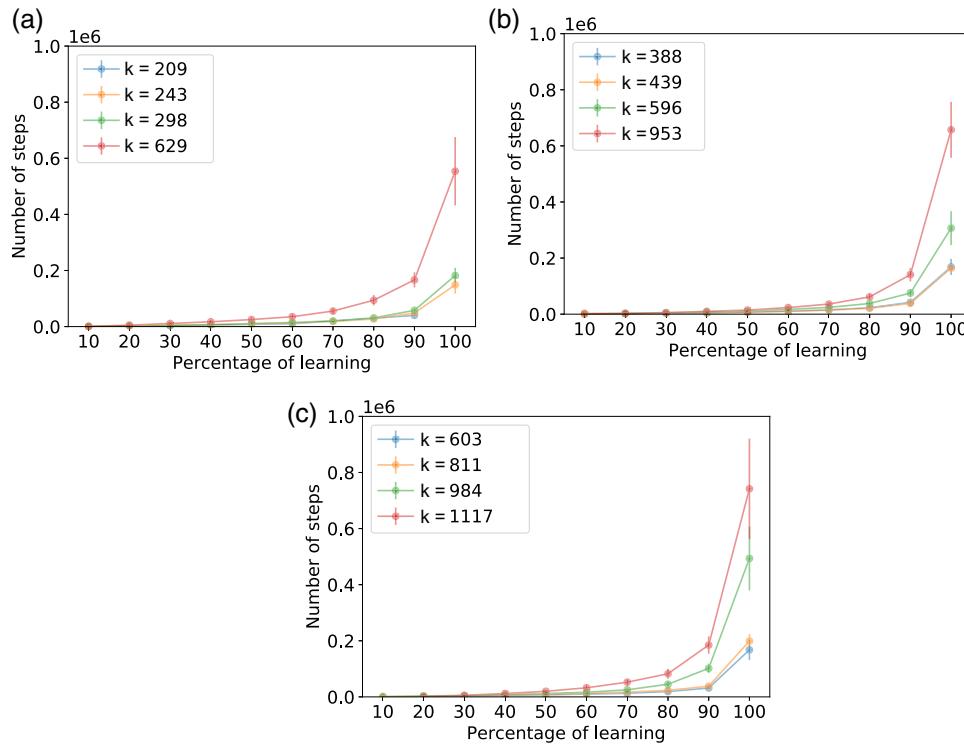


FIG. 7. Curves of efficiency obtained by selecting the hubs as starting nodes in BA networks for the Look-ahead dynamics. The legends indicate the node degree of the starting node.  $m$  is the number of links established by the node added in each step of the BA algorithm. (a)  $m = 3$ , (b)  $m = 4$ , and (c)  $m = 5$ .

## IV. RESULTS AND DISCUSSION

The results in this section all refer to the number of steps that had to be taken in order to discover a given percentage of the network. The total number of steps is considered to be the sum of steps of all the walks. This was done to emulate the cost of acquiring knowledge about the network.

The experiments were performed considering 100 different respective starting points in each network. The results shown in Secs. IV A–E are averaged over realizations starting from each one of the 100 nodes. First, the nodes are ordered according to their degrees. Next, the selected starting points are taken every 50 nodes in this sequence, implying 2% sampling of the nodes. This was done to better understand the influence of the centrality in the network on the total number of steps necessary to discover the whole network.

### A. ER network discovery

The studied ER networks are composed of 5000 nodes with average degree  $\langle k \rangle$  varying from 3 to 30. In Fig. 3, we show the number of steps required for covering a given percentage of the network, with respect to the three considered dynamics. A prototypical pattern can be identified in all cases, the only difference being an overall decrease with the average degree of the network.

Interestingly, there are no major differences between the extended self-avoiding and look ahead self avoiding types of walks. Avoiding or not the neighborhood of the walk bears little influence on the efficiency.

### B. Other models

When compared to other models we notice that the patterns displayed on ER networks are also present on almost all other models, this can be seen in Fig. 4.

Some interesting behaviors have been identified. First, almost every considered network model and walk model combinations present the same decreasing pattern observed for the ER networks. The most deviating curves refer to the BA and CM models, which present a significant standard deviation on look-ahead walks. WS networks (see Fig. 8) also exhibit a slight variance in efficiency, especially for the extended walk. Such small changes are probably caused by the existence of a few shortest paths compared to the regular lattice (for this model, we chose to have only 3% of the edges being rewired). This effect is more pronounced in the extended dynamics because it combines the two limitations from the other considered methods: it takes paths that are analogous to the normal dynamics, making it harder to reach the distant part of the network, and it always checks the neighborhood of the walk, causing the nodes in the vicinity of the brain to be checked several times.

### C. Influence of the starting node

We also studied the influence of the centrality of the chosen starting node on the total number of steps required for exploring the whole network. We found that the centrality (computed via degree, betweenness, eigenvector centrality, and accessibility<sup>40</sup>) does not have a great influence on the final results, with the only exception being the BA networks when considering the look-ahead self avoiding walks (see Figs. 5 and 6). Notably, this influence is only significant when the node of origin is a hub. In order to understand if this is a property specific to BA networks, we compared this model with CM networks with the same degree distribution. The same behavior was observed, leading us to believe that the actual degree distribution has a more important role in this dynamics than the model itself.

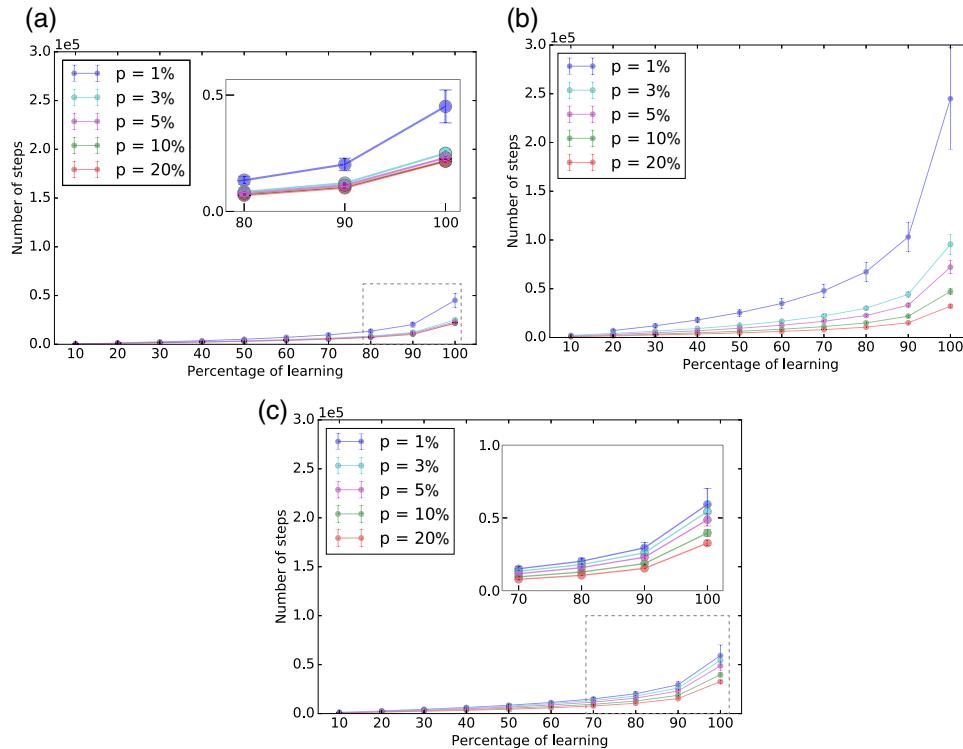


FIG. 8. Efficiency curves obtained for WS networks with varying rewiring probability  $p$  and  $\langle k \rangle = 4$ . (a) WS Standard, (b) WS Extended, and (c) WS Look-Ahead.

By selecting only the four nodes with the greatest degrees for each network, we observe that the larger the degree, the less efficient is the discovery, as illustrated in Fig. 7. This can be explained by the relatively high number of walks that start at the hub and terminate quickly. In the look-ahead dynamics, starting a walk at a hub tends to be highly inefficient since the agent checks every node around the hub. The number of these nodes typically surpasses 10% of the network. Moreover, since the known nodes are avoided, the chance of hitting a dead end is very high. This phenomenon can be seen in the video provided in the [supplementary material](#).

#### D. Structure influence

To analyze the effects of the structure on the efficiency of the discovery process, we studied WS networks with varying probabilities  $p$  of rewiring (see Fig. 8). We can see that

the extended self-avoiding walk has a much lower efficiency, which is mainly because the neighborhood of a walk is highly overlapping due to the high clustering coefficient. Because of such a high local clustering observed in each step, the nodes around the starting node are counted several times. The standard walk does not pose such a problem since it does not discover the nodes' neighbors and the look-ahead avoids known neighborhood thus generating very little overlap.

#### E. Real networks

The experiment using the WOS real-world network resulted in efficiency curves resembling those obtained for the BA and CM networks, as shown in Fig. 9. The selection of hubs as starting nodes was also found to play a major role in the dynamics. This corroborates the idea that the influence of

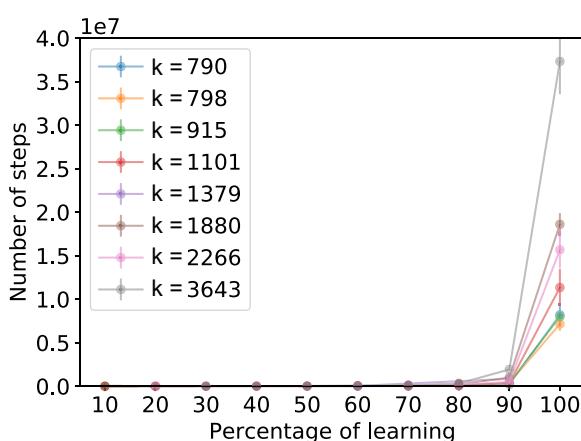


FIG. 9. Learning curves obtained by simulations of the extended dynamics on the considered real-world network. The legends indicate the node degree of the starting node.

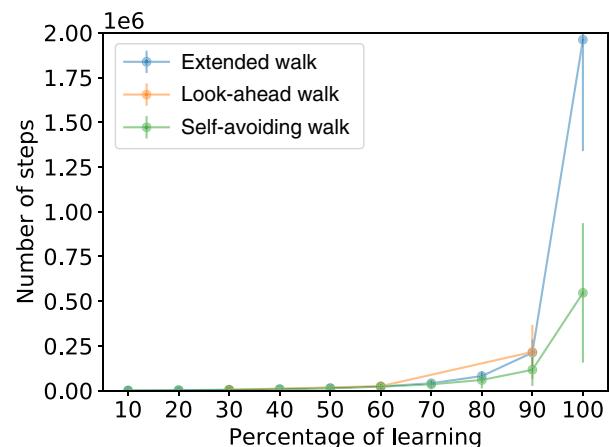


FIG. 10. Learning curves obtained for the CAL network. The best performance was achieved for the self-avoiding random walk, when the full coverage is considered.

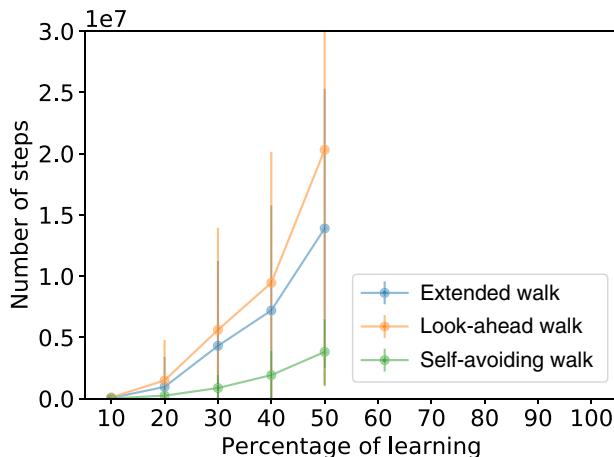


FIG. 11. Learning curves obtained for the PG network. The best performance was achieved for the self-avoiding random walk.

the starting node on the dynamics is strongly related to its degree and that this behavior does not seem to depend on other, more complex, topological characteristics, which are present in the WOS network.

Considering the California WWW network, CAL (see Fig. 10), the best performance was obtained with the traditional self-avoiding random walk. However, all three strategies displayed a similar behavior when we consider the number of steps taken to acquire a smaller portion of the network. The effectiveness of the self-avoiding random walk can also be observed in the power grid network (PG), as shown in Fig. 11. Even when a small percentage of the network coverage is considered, the self-avoiding random walk outperforms the other considered walks. We have not considered the full coverage of the network in this case because of the high computational cost associated with the exploration of the whole power grid network.

## V. CONCLUSIONS

Better understanding how knowledge is acquired in several circumstances is of fundamental importance in science and technology, as it allows predictions to be made about the behavior of the studied systems, as well as contributes to make those processes faster and more accurate. Being a particularly interesting and important type of knowledge acquisition, self-learning underlies several real-world systems, including human consciousness, risk assessment and prognostic, prospection, as well as the formation of social communities. Indeed, to a good extent, the rise of the WWW can be understood as a kind of self-knowledge.

While a good deal of works have investigated knowledge acquisition (see, e.g., Refs. 11 and 41–46), self-knowledge has received substantially less attention.<sup>47</sup> The present work approached this subject by using powerful concepts derived from network science. In particular, we investigated the self-knowledge acquisition as being a process derived from a random walk dynamics taking place in an information network. We considered three different types of dynamics and a diverse set of complex network models, as well as a real-world network representing the knowledge structure of a scientific

area. By considering different combinations of network models and dynamics, three aspects were investigated in terms of acquisition efficiency: dependence on the dynamics, on the network type, and on the local characteristics of the starting point.

The main contribution of this study is the result that the efficiency of self-knowledge acquisition, at least as approached from the point of view of the speed in which this dynamics unfolds and for the considered configurations and choices, hardly depends on either topology or the learning strategy considered. This is a surprising result itself, as it hints that the several paths to self-knowledge are similar regarding their speed efficacy.

Similar results, though slightly more dependent on topology and dynamics, have been obtained in a previous investigation,<sup>11</sup> but the extension of this principle to the here-reported self-knowledge is even more surprising because, unlike in that work, now the exploratory incursions are always performed with a fixed headquarter, so that the topology surrounding that node would be expected to have a greater impact on the self-knowledge acquisition.

As for future works, we intend to conduct a systematic study of the influence of topology, dynamics, and starting node in a large variety of complex systems, including social,<sup>48</sup> information,<sup>49–51</sup> semantic,<sup>52–54</sup> and technological networks.<sup>55</sup> More specifically, we intend to extend the analysis performed here for a broader range of random walks<sup>22,56</sup> and centrality measurements to select the starting nodes.<sup>57</sup> Future investigations should also study whether some of the adopted models can reproduce the dynamics of acquisition and transmission of knowledge.

## SUPPLEMENTARY MATERIAL

See the [supplementary material](#) for the illustration of network exploration dynamics obtained for the adopted random walks in a small network.

## ACKNOWLEDGMENTS

T.S.L. thanks Capes-Brazil for sponsorship. H.F.A. acknowledges Capes-Brazil for sponsorship. F.N.S. thanks FAPESP (Grant Nos. 15/08003-4 and 17/09280-7) for sponsorship. C.H.C thanks FAPESP (Grant No. 15/18942-8) for financial support. D.R.A. acknowledges FAPESP (Grant Nos. 16/19069-9 and 17/13464-6) for financial support. L.F.C. thanks CNPq (Grant No. 307333/2013-2) and NAP-PRP-USP for sponsorship. This work has been supported also by FAPESP (Grant Nos. 11/50761-2 and 15/22308-2).

<sup>1</sup>T. Stonier, “Information as a basic property of the universe,” *Biosystems* **38**, 135–140 (1996).

<sup>2</sup>R. Pastor-Satorras and A. Vespignani, “Epidemic spreading in scale-free networks,” *Phys. Rev. Lett.* **86**, 3200 (2001).

<sup>3</sup>Y. Moreno, M. Nekovee, and A. F. Pacheco, “Dynamics of rumor spreading in complex networks,” *Phys. Rev. E* **69**, 066130 (2004).

<sup>4</sup>R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, “Epidemic processes in complex networks,” *Rev. Mod. Phys.* **87**, 925 (2015).

<sup>5</sup>B. Danila, Y. Yu, J. A. Marsh, and K. E. Bassler, “Optimal transport on complex networks,” *Phys. Rev. E* **74**, 046106 (2006).

<sup>6</sup>M. Markosova, “Network model of human language,” *Phys. A: Stat. Mech. Appl.* **387**, 661–666 (2008).

<sup>7</sup>D. R. Amancio, "Probing the topological properties of complex networks modeling short written texts," *PLoS ONE* **10**, e0118394 (2015).

<sup>8</sup>H. F. de Arruda, L. da F. Costa, and D. R. Amancio, "Topic segmentation via community detection in complex networks," *Chaos: Interdiscip. J. Nonlinear Sci.* **26**, 063120 (2016).

<sup>9</sup>E. Bullmore and O. Sporns, "Complex brain networks: Graph theoretical analysis of structural and functional systems," *Nat. Rev. Neurosci.* **10**, 186–198 (2009).

<sup>10</sup>H. F. de Arruda, F. N. Silva, C. H. Comin, D. R. Amancio, and L. d. F. Costa, "Connecting network science and information theory," preprint [arXiv:1704.03091v2](https://arxiv.org/abs/1704.03091v2) (2017).

<sup>11</sup>H. F. de Arruda, F. N. Silva, L. da F. Costa, and D. R. Amancio, "Knowledge acquisition: A complex networks approach," *Inf. Sci.* **421**, 154–166 (2017).

<sup>12</sup>G. Yan, T. Zhou, B. Hu, Z.-Q. Fu, and B.-H. Wang, "Efficient routing on complex networks," *Phys. Rev. E* **73**, 046108 (2006).

<sup>13</sup>M. Boguna, D. Krioukov, and K. C. Claffy, "Navigability of complex networks," *Nat. Phys.* **5**, 74 (2009).

<sup>14</sup>S. Lee, S.-H. Yook, and Y. Kim, "Searching method through biased random walks on complex networks," *Phys. Rev. E* **80**, 017102 (2009).

<sup>15</sup>Y. Kim, S. Park, and S.-H. Yook, "Network exploration using true self-avoiding walks," *Phys. Rev. E* **94**, 042309 (2016).

<sup>16</sup>C. Castellano, S. Fortunato, and V. Loreto, "Statistical physics of social dynamics," *Rev. Mod. Phys.* **81**, 591 (2009).

<sup>17</sup>N. J. McCullen, A. M. Rucklidge, C. S. Bale, T. J. Foxon, and W. F. Gale, "Multiparameter models of innovation diffusion on complex networks," *SIAM. J. Appl. Dyn. Syst.* **12**, 515–532 (2013).

<sup>18</sup>K. Pearson, "The problem of the random walk," *Nature* **72**, 342 (1905).

<sup>19</sup>L. Rayleigh, "The problem of the random walk," *Nature* **72**, 318 (1905).

<sup>20</sup>E. Agirre and A. Soroa, "Personalizing pagerank for word sense disambiguation," in *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics* (Association for Computational Linguistics, 2009), pp. 33–41.

<sup>21</sup>M. Rosvall and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure," *Proc. Natl. Acad. Sci.* **105**, 1118–1123 (2008).

<sup>22</sup>N. Masuda, M. A. Porter, and R. Lambiotte, "Random walks and diffusion on networks," *Phys. Rep.* **716–717**, 1–58 (2017).

<sup>23</sup>D. R. Amancio, "Comparing the topological properties of real and artificially generated scientific manuscripts," *Scientometrics* **105**, 1763–1779 (2015).

<sup>24</sup>N. Craswell and M. Szummer, "Random walks on the click graph," in *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (ACM, 2007), pp. 239–246.

<sup>25</sup>L. Grady, "Random walks for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **28**, 1768–1783 (2006).

<sup>26</sup>N. Duhan, A. Sharma, and K. K. Bhatia, "Page ranking algorithms: A survey," in *IEEE International Advance Computing Conference, 2009. IACC 2009* (IEEE, 2009), pp. 1530–1537.

<sup>27</sup>L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web." Technical Report (Stanford InfoLab, 1999).

<sup>28</sup>U. Feige, "A tight upper bound on the cover time for random walks on graphs," *Random Struct. Algorithms* **6**, 51–54 (1995).

<sup>29</sup>P. Erdos and A. Rényi, "On the evolution of random graphs," *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17–60 (1960).

<sup>30</sup>A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science* **286**, 509–512 (1999).

<sup>31</sup>M. E. Newman, "The structure and function of complex networks," *SIAM Rev.* **45**, 167–256 (2003).

<sup>32</sup>D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks," *Nature* **393**, 440–442 (1998).

<sup>33</sup>B. M. Waxman, "Routing of multipoint connections," *IEEE J. Sel. Areas Commun.* **6**, 1617–1622 (1988).

<sup>34</sup>V. Batagelj, A. Mrvar, A. Ferligoj, and P. Doreian, "Generalized blockmodeling with Pajek," *Metodoloski zvezki* **1**, 455 (2004).

<sup>35</sup>P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," *Soc. Networks* **5**, 109–137 (1983).

<sup>36</sup>See See <http://www.webofscience.com> for access to the dataset of papers and respective references.

<sup>37</sup>J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *J. ACM* **46**, 604–632 (1999).

<sup>38</sup>A. N. Langville and C. D. Meyer, "A reordering for the pagerank problem," *SIAM J. Sci. Comput.* **27**, 2112–2120 (2006).

<sup>39</sup>F. N. Silva, D. R. Amancio, M. Bardosova, L. da F. Costa, and O. N. Oliveira Jr, "Using network science and text analytics to produce surveys in a scientific topic," *J. Informetr.* **10**, 487–502 (2016).

<sup>40</sup>B. A. N. Travençolo and L. d. F. Costa, "Accessibility in complex networks," *Phys. Lett. A* **373**, 89–95 (2008).

<sup>41</sup>I. T. Koponen and M. Nousiainen, "Modelling students' knowledge organisation: Genealogical conceptual networks," *Phys. A: Stat. Mech. Appl.* **495**, 405–417 (2018).

<sup>42</sup>I. T. Koponen and M. Nousiainen, "Concept networks in learning: Finding key concepts in learners' representations of the interlinked structure of scientific knowledge," *J. Complex Netw.* **2**, 187–202 (2014).

<sup>43</sup>I. Iacopini, S. Milojevic, and V. Latora, "Network dynamics of innovation processes," *Phys. Rev. Lett.* **120**, 048301 (2018).

<sup>44</sup>M. Lin and N. Li, "Scale-free network provides an optimal pattern for knowledge transfer," *Phys. A: Stat. Mech. Appl.* **389**, 473–480 (2010).

<sup>45</sup>B. Cao, S. hua Han, and Z. Jin, "Modeling of knowledge transmission by considering the level of forgetfulness in complex networks," *Phys. A: Stat. Mech. Appl.* **451**, 277–287 (2016).

<sup>46</sup>H.-M. Zhu, S.-T. Zhang, and Z. Jin, "The effects of online social networks on tacit knowledge transmission," *Phys. A: Stat. Mech. Appl.* **441**, 192–198 (2016).

<sup>47</sup>H. Wang, J. Wang, L. Ding, and W. Wei, "Knowledge transmission model with consideration of self-learning mechanism in complex networks," *Appl. Math. Comput.* **304**, 83–92 (2017).

<sup>48</sup>X. Chen, X. Xiong, M. Zhang, and W. Li, "Public authority control strategy for opinion evolution in social networks," *Chaos: Interdiscip. J. Nonlinear Sci.* **26**, 083105 (2016).

<sup>49</sup>L. Yin, L.-C. Shi, J.-Y. Zhao, S.-Y. Du, W.-B. Xie, F. Yuan, and D.-B. Chen, "Heterogeneous information network model for equipment-standard system," *Phys. A: Stat. Mech. Appl.* **490**, 935–943 (2018).

<sup>50</sup>H.-L. Zhou and X.-D. Zhang, "Dynamic robustness of knowledge collaboration network of open source product development community," *Phys. A: Stat. Mech. Appl.* **490**, 601–612 (2018).

<sup>51</sup>F. Breve, L. Zhao, M. Quiles, W. Pedrycz, and J. Liu, "Particle competition and cooperation in networks for semi-supervised learning," *IEEE Trans. Knowl. Data Eng.* **24**, 1686–1698 (2012).

<sup>52</sup>D. R. Amancio, O. N. Oliveira Jr., and L. da F. Costa, "Structure-semantics interplay in complex networks and its effects on the predictability of similarity in texts," *Phys. A: Stat. Mech. Appl.* **391**, 4406–4419 (2012).

<sup>53</sup>D. R. Amancio, "A complex network approach to stylometry," *PLoS ONE* **10**, e0136076 (2015).

<sup>54</sup>H. Chen, X. Chen, and H. Liu, "How does language change as a lexical network? An investigation based on written Chinese word co-occurrence networks," *PLoS ONE* **13**, e0192545 (2018).

<sup>55</sup>K. C. Chen, M. Chiang, and H. V. Poor, "From technological networks to social networks," *IEEE J. Sel. Areas Commun.* **31**, 548–572 (2013).

<sup>56</sup>R. Pemantle *et al.*, "A survey of random processes with reinforcement," *Probab. Surv.* **4**, 1–79 (2007).

<sup>57</sup>M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, "Identification of influential spreaders in complex networks," *Nat. Phys.* **6**, 888 (2010).