

## Exposure theory for learning complex networks with random walks

ANDREI A. KLISHIN

*Department of Bioengineering, University of Pennsylvania, 210 S. 33rd Street, 240 Skirkanich Hall, Philadelphia, PA 19104-6321, USA*

AND

DANI S. BASSETT<sup>†</sup>

*Department of Bioengineering, Department of Physics & Astronomy, Department of Electrical & Systems Engineering, Department of Neurology and Department of Psychiatry, University of Pennsylvania, 210 S. 33rd Street, 240 Skirkanich Hall, Philadelphia, PA 19104-6321, USA and Santa Fe Institute, Santa Fe, NM 87501, USA*

<sup>†</sup>Corresponding author. Email: [dsb@seas.upenn.edu](mailto:dsb@seas.upenn.edu)

[Received on 18 April 2022; editorial decision on 18 April 2022; accepted on 16 May 2022]

Random walks are a common model for the exploration and discovery of complex networks. While numerous algorithms have been proposed to map out an unknown network, a complementary question arises: in a known network, which nodes and edges are most likely to be discovered by a random walker in finite time? Here, we introduce exposure theory, a statistical mechanics framework that predicts the learning of nodes and edges across several types of networks, including weighted and temporal, and show that edge learning follows a universal trajectory. While the learning of individual nodes and edges is noisy, exposure theory produces a highly accurate prediction of aggregate exploration statistics.

*Keywords:* graph learning; complex networks; random walks.

### 1. Introduction

Random walks are a common baseline model of dynamical processes on networks, representing phenomena from search to disease spreading to communication [1]. The problem of random walk statistics, such as the number of visited nodes and the return probability, was first formulated and solved on infinite, periodic, low-dimensional lattices [2, 3], but the early numerical studies on finite, complex, irregular networks revealed quite different exploration behaviours [4–6]. Along with node- and edge-level exploration statistics, many formal results have been obtained for aggregate metrics, such as cover and return times [7–10]. Random walk statistics can be used to infer the underlying network structure, from node centrality [11] to community structure [12, 13]. While simple random walks independently revisit known nodes and edges, a variety of modified random walk algorithms have been developed to speed up network exploration or deliberately revisit the known parts [4, 6, 11, 14–16].

Beyond their relevance for dynamical network processes in the environment, random walks are also relevant for cognitive processes taking place in the human mind. Recent experiments show that humans (and machines) can learn network structures from observing random walks taken upon them [17–20]. Unlike machines, humans commit mental errors in learning networks, which serves to foreground large-scale network structure over fine-grained details [21]. The dual needs to minimize the impact of such errors while maximizing the communication efficiency might serve as selection pressures on communication network architectures [22].

Our goal is to understand how the learning of complex networks by random walks is driven by the interplay of the finite length of random walks and the structure of the learned networks. In order to quantify this interplay, in this article, we propose and validate *exposure theory*, a framework similar to the conventional Boltzmann statistical mechanics in that it predicts a local probability of visiting an individual edge and provides a rule for combining such probabilities to predict aggregate statistics. We validate both local and aggregate predictions against direct stochastic random walk simulations and find exposure theory to be highly accurate at a fraction of the computational cost. Exposure theory systematizes and generalizes a variety of previous results on random walks and can be further expanded to account for human information processing.

## 2. Mathematical formalism

We consider random walks on an undirected, weighted, time-dependent network described by the adjacency matrix  $\mathbf{A}(t)$ . The *learning* of the network is described by the integer-valued memory matrix  $\mathbf{M}$ : every time the walk traverses (visits) the edge  $(i, j)$ , we add 1 to the element  $M_{ij}$ . How are the different elements of  $\mathbf{M}$  populated over a finite time? On one side, we simulate several random walk realizations and compute the statistics. On the other side, we introduce the exposure ensemble, that predicts the memory distribution directly, but approximately.

We first compute the steady-state probability of visiting a particular edge  $p_{ij}$ . In random walks on a weighted network, the conditional probability  $P(j|i, t)$  of the step  $i \rightarrow j$  is given by the row-normalized adjacency matrix. We combine it with the steady-state probability of ending up on a particular node  $\pi_i(t)$  so that the row norm cancels out:

$$P(j|i, t) = \frac{A_{ij}(t)}{\sum_j A_{ij}(t)}; \quad \pi_i(t) = \frac{\sum_j A_{ij}(t)}{\sum_{ij} A_{ij}(t)}; \quad (2.1)$$

$$p_{ij}(t) = \pi_i(t)P(j|i, t) = \frac{A_{ij}(t)}{\sum_{ij} A_{ij}(t)}. \quad (2.2)$$

What is the probability of visiting the edge  $(i, j)$  multiple times and thus getting multiple memory counts  $M_{ij}$ ? Technically, subsequent visitations of the same edge are not conditionally independent. The conditional dependence is especially strong in lattices [2], in Watts–Strogatz networks derived from one-dimensional lattices [23], and other networks embedded in a low-dimensional space [10] (see Appendix G for an explicit test). However, many other complex networks do not have a low-dimensional latent space and thus have very short mixing or relaxation times [10]. For such networks, we can build exposure theory resting on three key assumptions: subsequent edge visitations are conditionally independent, their probability distribution follows the instantaneous steady-state  $p_{ij}(t)$  and  $p_{ij} \ll 1$ . In this case, the accumulation of memories of each edge follows an independent, possibly non-stationary Poisson process (see Fig. 1 and Appendix A):

$$M_{ij}(t) \sim \text{Pois}(E_{ij}(t)); \quad E_{ij}(t) \equiv \sum_1^t p_{ij}(t'), \quad (2.3)$$

where we defined the quantity  $E_{ij}$  as the *integral exposure* of the edge  $(i, j)$ . The name is inspired by the operation of a camera: the shutter is opened for a specific time  $t$ , during which the film or sensor inside integrates the incoming flux of light. The entries  $E_{ij}$  form a matrix that depends on the length of the random walk.

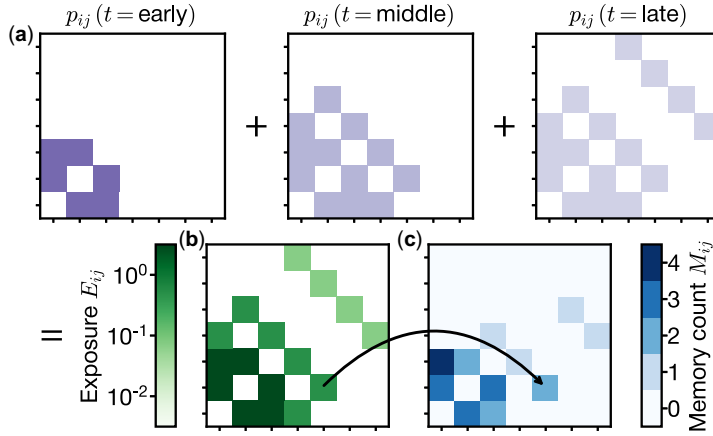


FIG. 1. Exposure theory consists of a deterministic accumulation of exposure and a stochastic draw of memory count. (a) As the network grows, at every time point  $t$  we compute the matrix of edge visit probabilities  $\mathbf{p}(t)$  and sum them up to obtain the integral exposure matrix  $\mathbf{E}$  (b). (c) Each entry of the exposure matrix is the parameter of an independent Poisson distribution from which we draw the integer entries of the memory count matrix  $\mathbf{M}$ .

How does the exposure matrix relate to properties of a network, such as its weight distribution and time dependence? We first consider a time-independent network, in which the visitation probability is also time independent. In this case, the sum in (2.3) consists of  $t$  identical terms and is trivially computed to yield:

$$E_{ij}(t) = t \cdot \mathcal{E}_{ij}; \quad \mathcal{E}_{ij} \equiv p_{ij} = \frac{A_{ij}}{\sum_{ij} A_{ij}}, \quad (2.4)$$

where  $\mathcal{E}_{ij}$  is the *specific exposure* (per random walk step). Thus for time-independent networks, the exposure of each edge grows linearly in time in proportion to the relative edge weight. For unweighted networks, the exposure of all edges is the same  $E_{ij} = t/m$ , where  $m$  is the number of edges.

In contrast, if the underlying network is time-dependent, then we assume that it evolves under some deterministic protocol  $A_{ij}(\tau)$  at the same time as the random walk unfolds. We synchronize the random walk time  $t$  with the evolution time  $\tau$  by setting their ratio to  $t/\tau \equiv D$  and refer to this quantity as the *dilation*. A low value of the dilation parameter indicates that network evolution is fast compared to the random walk, whereas a high value indicates that the random walk can explore the network before it changes significantly. In this case, the exposure can be computed as follows:

$$E_{ij}(t) = D \cdot \mathcal{E}_{ij}(t/D); \quad \mathcal{E}_{ij}(\tau) \equiv \int_0^\tau p_{ij}(\tau') d\tau', \quad (2.5)$$

where  $\mathcal{E}_{ij}(\tau)$  is the time-dependent version of specific exposure. Varying the value of dilation  $D$  thus rescales the absolute value of integral exposure, but stretches out its evolution.

The specific exposure matrix  $\mathcal{E}$  is another way to represent a network with a matrix, alongside others such as adjacency, Laplacian and modularity matrices [24]. While other matrices do not change their meaning significantly when all elements are uniformly rescaled by a scalar, such rescaling is crucial

in converting from specific to integral exposure. The integral exposure matrix describes the *canonical ensemble* of random walk memories over the network since the total number of memory counts fluctuates around the average of  $t$  (see Appendix B). In a conventional Boltzmann canonical ensemble the probability of a state  $p(s) \propto e^{-\beta H(s)}$  is driven by the product of the inverse temperature  $\beta$  and the Hamiltonian  $H(s)$ ; in the exposure ensemble the memory distribution (2.3) is driven by the product of the prefactor  $t$  or  $D$  and the edge-specific exposure  $\mathcal{E}_{ij}$ . While the specific probability formula differs, a range of computational techniques carry over, especially for coarse graining (see Appendix B).

Similarly to the Boltzmann ensemble, we are interested in the limiting cases of large and small exposure. The large exposure limit is easily reached by making the prefactor  $t$  or  $D$  large. In this case, the Poisson distribution is well-approximated by its mean, and thus the number of counts of each edge is nearly equal to its exposure. The small exposure limit is more subtle. The probability of having more than zero memories of a particular edge is given by the expression:

$$P(M_{ij} > 0) = 1 - e^{-E_{ij}(t)}, \quad (2.6)$$

which implies that the probability of learning any edge *locally* follows a universal exponential curve in terms of its exposure.

For small learning times, the learning of individual edges remains highly noisy. However, since the learning of each edge is independent, we expect to get much more accurate results by aggregating across multiple edges. In particular, exposure theory can predict the *fraction of edges* that have been learned:

$$V(t) = 1 - \frac{1}{m} \sum_{ij \in A_{ij}} e^{-E_{ij}(t)} \leq 1 - e^{-t/m}, \quad (2.7)$$

where the sum runs over the edges that exist in the adjacency matrix. The bound comes from applying Jensen's inequality to the sum (see Appendix C). The right-hand side expression corresponds to the fraction of edges learned in an unweighted network of  $m$  directed edges. Therefore, Jensen's inequality imposes a fundamental limit on the speed of learning, regardless of the distribution of network weights and the time-dependent protocol for network change. The predictions of both the detailed shape of  $V(t)$  and the Jensen bound can also be tested by direct comparison to stochastic random walk simulations.

### 3. Validation

Exposure theory has yielded two predictions: a local learning curve (2.6) and an aggregate exploration curve (2.7). While these predictions assume conditional independence of visitations, we validate them with stochastic simulations that faithfully capture the conditional dependence on two static networks and one dynamic network. The first network (Copperfield) tracks the co-occurrence of nouns and adjectives in the text of *David Copperfield*, a novel by Charles Dickens, and is unweighted and undirected [25]. The second network (US airports) tracks the passenger flow between the top-500 busiest airports in the USA [26]. The passenger flow varies by a factor of  $10^5$  across the edges, but is symmetric within each pair, resulting in a weighted undirected network.

On each static network, we simulate 100 replicas of stochastic random walk realizations that begin at a randomly chosen node. In order to test the local prediction, for each edge, we compute the fraction of replicas in which it was visited by time  $t$ . We convert time  $t$  into integral exposure by rescaling it: for the Copperfield network, the rescaling is uniform for each edge  $t \rightarrow t/m$ , whereas for the US airports network the rescaling is edge-specific  $t \rightarrow t \cdot \mathcal{E}_{ij}$ . After rescaling, we average the probability across the

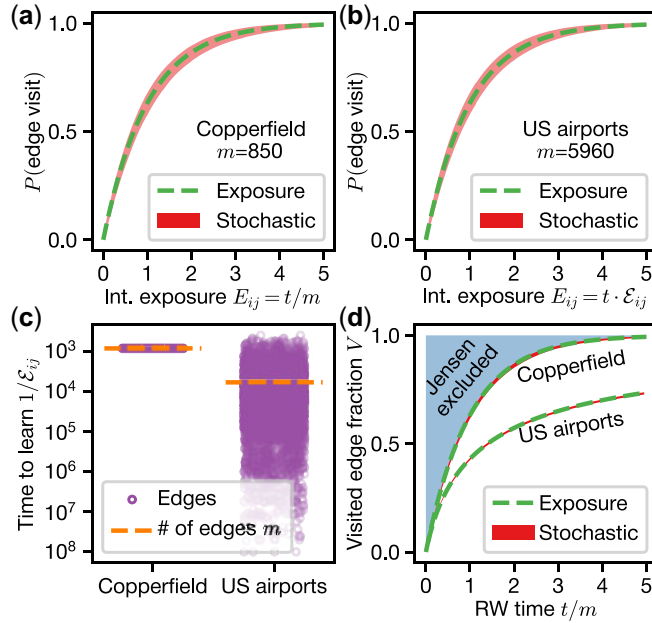


FIG. 2. Validation of local and aggregate exposure theory predictions for time-independent networks. (a–b) Universal curve of edge visit probability (dashed curve) describes the stochastic probability over 100 runs (shaded area, mean $\pm$ std) averaged over  $m$  directed edges. Time is scaled by edge exposure: uniform for the unweighted network (a) and edge-specific for the weighted network (b). (c) Scatterplot of time to learn each edge in the unweighted and weighted networks. (d) Visited edge fraction against time scaled by the number of edges. The shaded region is excluded by the Jensen bound, which the unweighted network saturates.

edges in each network. In both cases (Fig. 2(a,b)) the prediction of (2.6) lines up with the stochastic results extremely well, which points to the universal nature of edge learning and several practical corollaries. First, learning edges in the pair  $(i,j) - (j,i)$  is equally likely but independent. Second, in unweighted networks such as Copperfield, learning all edges is equally likely. Third, the characteristic time to learn any edge is  $1/\epsilon_{ij}$ , which varies over many orders of magnitude (Fig. 2(c)).

After the local predictions, we test the aggregate prediction of the visited edge fraction  $V(t)$ . The learning of the unweighted Copperfield network exactly saturates the Jensen bound, while the weighted US airports network significantly underperforms (Fig. 2(d)). In both cases, the variance of  $V$  is extremely small, which points to the accuracy of exposure theory predictions of aggregate metrics. Similar functional forms of the aggregate learning curve have been previously observed or assumed without a derivation [4, 6, 15].

Having confirmed the universal learning curve on static networks, we extend exposure theory predictions to a temporal network that we adapted from Ref. [27]. The authors represented several popular linear algebra textbooks as temporal networks, where nodes are key concepts and edges are co-occurrence of concepts in a sentence. The time-dependent adjacency matrix is defined as  $A_{ij}(\tau) = A_{ij} \cdot [F_{ij} < \tau]$ , where  $A_{ij}$  is the number of co-occurrences of the pair  $(i,j)$ ,  $F_{ij}$  is the filtration order (time of first co-occurrence),  $[\cdot]$  is an indicator function, and time  $\tau \in [0, \tau_{\max}]$  is measured in sentences of the book. Out of the 10 textbooks, here we analyse the network of the textbook by Treil (Fig. 3(a)).

We compare the taught and the learned networks by three architectural metrics: edge weight  $A$ , filtration order  $F$ , and core-periphery (CP) structure. The cumulative distribution functions (CDFs) for the learned edges are biased towards higher  $A$  and lower  $F$  with respect to the textbook, but the bias

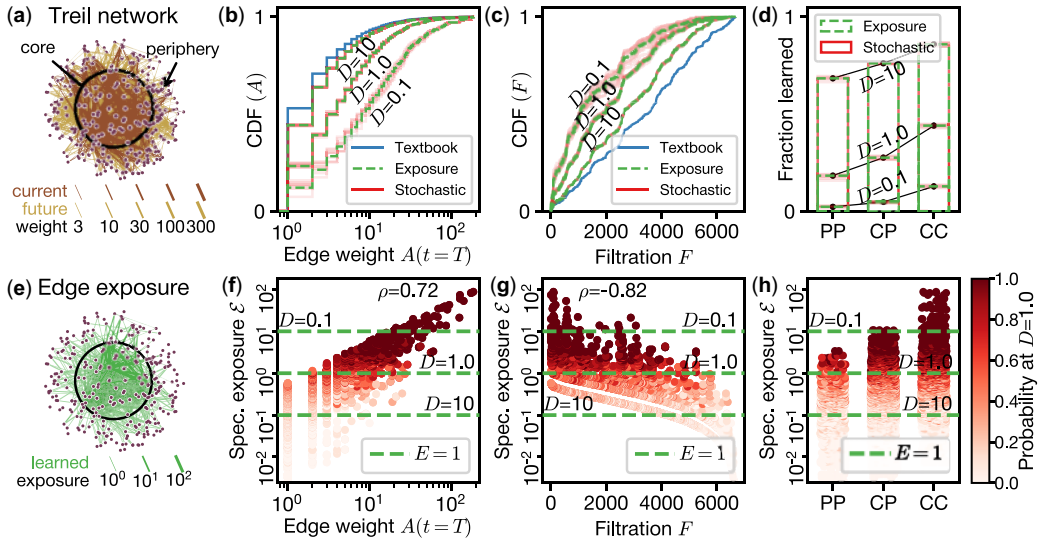


FIG. 3. Exposure theory predicts the architecture of learned networks. (a) The Treil network edges vary widely in both edge weight and filtration order, and together form a robust core-periphery structure. (b–c) Cumulative distribution functions (CDFs) for edge weight by the end of the book ( $A(t = T)$ , b) and filtration order ( $F$ , c) for the textbook network (solid curve), stochastic simulations (transparent curves), and exposure prediction (dashed curve). (d) Fraction of edges learned in different structural positions: within-core (CC), core-periphery (CP), and within-periphery (PP). (e) The Treil network with edges valued by exposure. (f–h) Scatter plots of all edges in the network, with the shade corresponding to the probability of the edge being learned across 10 replicas at  $D = 1.0$ . The horizontal dashed lines show the condition  $E = D\mathcal{E} = 1$ , which is a good separator of learned and non-learned edges. Panels (f,g) include the Spearman correlation coefficient  $\rho$  between specific exposure  $\mathcal{E}$  and edge weight  $A(t = T)$  and filtration  $F$ , respectively,  $\log_{10}(\rho) < -12$ .

decreases with dilation  $D$  (Fig. 3(b,c), see Appendix D). Edges within the core (CC) are most likely to be learned, followed by CP and within-periphery (PP) edges (Fig. 3(d)). In all cases exposure theory provides an excellent match to the stochastic CDFs.

Finer architectural details of learned networks are revealed through the lens of edge-level specific exposure (Fig. 3(e)). The specific exposure of individual edges  $\mathcal{E}_{ij}$  depends locally on the edge weight  $A_{ij}$  but also non-locally on the sum of all edge weights  $\sum_{ij} A_{ij}(t)$  and temporally on the filtration order  $F_{ij}$  (Spearman correlation in Fig. 3(f–g)). While edge weight  $A_{ij}$  for the Treil network spans two orders of magnitude, the range of specific exposure stretches out to nearly six orders of magnitude. As dilation  $D$  increases, more edges cross the exposure threshold  $E_{ij} = D\mathcal{E}_{ij} \simeq 1$  and are learned (Fig. 3(f–h)). Specific exposure predicts not only the binary learning outcome but also detailed statistics of the memory count  $M_{ij}$  at any dilation (see Appendix E).

#### 4. Discussion

The proposed exposure theory builds upon several previous results but simultaneously generalizes them under a single statistical mechanics framework. Exposure theory improves upon previous results in three aspects. First, we centre our analysis on the network edges and obtain results for nodes as a simple corollary via a coarse-graining process (see Appendix B). Second, we focus on the probability distributions and time dependence of quantities of interest as opposed to just the means. Third, we consider not just one



but all visitations of an edge or a node and thus do not need to introduce artificial absorbing states [1]. Exposure theory relies on seemingly stringent assumptions of network properties, but these assumptions hold for most real-world complex networks of non-spatial origin, as discussed in Refs. [10, 28] and as shown in the context of exposure predictions in Appendices E and F.

The two key predictions of exposure theory, the local visit probability ((2.6)) and the aggregate visit fraction ((2.7)), look deceptively similar mathematically but have been analysed separately before. The time to visit an edge or node is also commonly known as the first passage time, and the visitation probability in (2.6) serves as its CDF. Usually, an interest in the *mean* first passage time stimulates the development of computational methods [29], but Ref. [28] argued that the first passage time *distribution* has a universal exponential tail with a variable decay rate consistent with our (2.6). The decay rates of the first passage time distribution can be used to compute the cover time, or average time to learn the whole network [10]. The aggregate visit fraction in (2.7) tracks not only the time to visit all edges or nodes but also the detailed dynamics of approaching that limit, thus providing a theoretical backing to the exploration curves reported in Refs. [4, 6]. In addition to deriving the exploration curve, we provide a fundamental Jensen bound on the speed of exploration possible with random walks.

Random walks are a paradigmatic example of a dynamical process used to map out *a priori* unknown networks. However, the properties of such a dynamical process can produce a significantly distorted view of network architecture. For example, when the physical topology of the Internet is studied with the common traceroute algorithm, the measured network can have a spurious power-law node degree distribution even if the underlying network has narrowly distributed or even identical node degrees [30, 31]. Acknowledging this distortion led to the development of unbiased sampling techniques [32]. Exposure theory accounts for the distortions produced by the under-sampling bias of finite random walks (Fig. 3). While at the binary level (existence of nodes or edges), the sampling bias decreases and disappears with longer walk time, it remains to be seen if more subtle distortions of edge weight or node strength persist for longer times.

Apart from learning the detailed network topology, finite-time spreading processes such as random walks have been used to infer a variety of other structural metrics. The spreading time can serve as a natural multiscale lens to study network structure from community partitions [33, 34] to node centrality [35]. Finite-time spreading is also key to the quantum-like entanglement of network flows [36, 37]. Random walks inspire centrality measures for multiplex networks [38]. Hyperbolic network models allow tuning the network spectral dimension [39], which in turn drives the behaviour of dynamical processes such as synchronization and diffusion [40, 41]. Lastly, random walks are key to the diffusion part of reaction–diffusion systems [26, 42], which in turn are crucial to the temporal prediction of epidemic spreading and related processes [43, 44].

## 5. Conclusions

In this article, we studied random walk exploration of a broad class of complex networks that span from unweighted, undirected to weighted and temporal. For these networks, our proposed exposure theory produces detailed and accurate predictions from edge-level probability to aggregate learning curves, with no free parameters beyond those required to specify the random walk conditions. Exposure computations are orders of magnitude faster than stochastic simulations (see Appendix H for benchmarks). While the speed of exploration by simple random walks is limited by the Jensen bound, it can be surpassed by biased random walks [45], intelligent exploration algorithms [6] or multiple interacting walkers [15]. Exploration dynamics might also deviate from the universal curve for directed networks that support more complex steady states [4], or temporal networks that connect and disconnect components dynamically [46–48].

Even for simple random walks, the learned networks can be significantly distorted by the faults of human memory [21]; we are looking forward to extensions of exposure theory to explore such distortions in detail.

### Acknowledgements

The authors would like to thank N.H. Christianson for providing the linear algebra textbook network data, and C.W. Lynn and X. Xia for discussions about the modelling. The computational workflow in general and data management in particular for this work was primarily supported by the Signac data management framework [49, 50].

### Funding

The Army Research Office (DCIST-W911NF-17-2-0181 and GRAFTON-W911NF-16-1-0474 to D.S.B.); and the National Institutes of Mental Health (1-R21-MH-124121-01 to D.S.B.).

### Citation diversity statement

Recent work in several fields of science has identified a bias in citation practices such that papers from women and other minority scholars are under-cited relative to the number of such papers in the field [51–60]. Here, we sought to proactively consider choosing references that reflect the diversity of the field in thought, form of contribution, gender, race, ethnicity and other factors. First, we obtained the predicted gender of the first and last author of each reference by using databases that store the probability of a first name being carried by a woman [55, 61]. By this measure (and excluding self-citations to the first and last authors of our current article and references in this paragraph), our references contain 6.25% woman(first)/woman(last), 8.33% man/woman, 14.58% woman/man and 70.83% man/man. This method is limited in that (a) names, pronouns and social media profiles used to construct the databases may not, in every case, be indicative of gender identity and (b) it cannot account for intersex, non-binary or transgender people. Second, we obtained the predicted racial/ethnic category of the first and last author of each reference by databases that store the probability of a first and last name being carried by an author of colour [62, 63]. By this measure (and excluding self-citations and references in this paragraph), our references contain 10.9% author of colour (first)/author of colour (last), 20.54% white author/author of colour, 16.8% author of colour/white author and 51.76% white author/white author. This method is limited in that (a) names and Florida Voter Data to make the predictions may not be indicative of racial/ethnic identity and (b) it cannot account for Indigenous and mixed-race authors, or those who may face differential biases due to the ambiguous racialization or the ethnicization of their names. We look forward to future work that could help us to better understand how to support equitable practices in science.

### A. Derivation of exposure theory

Here, we provide a first-principles derivation of exposure theory from the three assumptions stated in the main text:

1. Subsequent edge visitations are conditionally independent from each other.
2. Probability distribution of visitation follows the instantaneous steady-state  $p_{ij}(t)$ .
3. Probability of visiting a particular edge in one step is small  $p_{ij}(t) \ll 1$ .

These assumptions are quite frequently satisfied for complex networks studied in this article. Assumptions 1 and 2 rely on fast mixing (short correlation time) of random walks, which we explicitly compute for



the studied networks in Section F. Assumption 1 additionally implies that the network is connected—otherwise the probability of visiting an edge in one connected component via a random walk from another component would be zero. Assumption 3 holds generally for any large network—for unweighted networks  $p_{ij} = 1/m \ll 1$  when there are many edges, while for weighted networks it holds so long as the weight of one or a few edges does not constitute a large fraction of the total weight of all edges.

As the accumulation of memories is stochastic, a random walk of length  $t$  may result in a distribution of possible memory matrices  $P(\mathbf{M}, t)$ . However, if the Assumptions 1 and 2 hold, the accumulation of memories of each edge is independent and follows a distribution  $P_{ij}(k, t)$ , where we used  $k \in \{0, 1, 2, \dots\}$  as an index of the distribution to simplify notation. Since, the only event that can happen to an edge is the addition of a count, the evolution of the distribution follows a relatively simple master equation:

$$\begin{aligned} P_{ij}(k, t+1) - P_{ij}(k, t) &= \begin{cases} p_{ij}(t) (P_{ij}(k-1, t) - P_{ij}(k, t)), & k > 0 \\ -p_{ij}(t) P_{ij}(k, t), & k = 0 \end{cases}, \end{aligned} \quad (\text{A.1})$$

$$P_{ij}(k, 0) = \delta_{k,0}, \quad (\text{A.2})$$

where the first equation expresses the dynamics of the distribution and the second one expresses the initial condition (the edge starts with no memories).

In general, the expression A.1 is an infinite system of coupled equations. However, we can attempt to solve them with the following ansatz:

$$P_{ij}(k, t) = \frac{(E_{ij}(t))^k e^{-E_{ij}(t)}}{k!}, \quad (\text{A.3})$$

which is the Poisson distribution with a single, yet-to-be-determined time-dependent parameter  $E_{ij}(t)$ . Substituting  $E_{ij} = 0$  recovers the initial condition of Eq. A.2. It remains for us to show that the Poisson distribution holds at all times and to find the growth law for  $E_{ij}$ .

We first simplify the right-hand side of Eq. A.1 as follows:

$$\begin{aligned} p_{ij}(t) (P_{ij}(k-1, t) - P_{ij}(k, t)) &= p_{ij}(t) P_{ij}(k, t) \left( \frac{k}{E_{ij}(t)} - 1 \right), \end{aligned} \quad (\text{A.4})$$

where we used the functional form of the Poisson distribution. We now observe that the whole expression is proportional to  $p_{ij} \ll 1$  by Assumption 3. Thus, the probability distribution cannot change too rapidly in a single step. We therefore approximate the finite difference on the left-hand side with a derivative:

$$\begin{aligned} P_{ij}(k, t+1) - P_{ij}(k, t) &\approx \Delta E_{ij} \frac{\partial P_{ij}(k, t)}{\partial E_{ij}} \\ &= \Delta E_{ij} P_{ij}(k, t) \left( \frac{k}{E_{ij}(t)} - 1 \right), \end{aligned} \quad (\text{A.5})$$

where we recovered an identical  $k$ - and  $t$ -dependent expression in the brackets that can be cancelled out.

From the transformed left- and right-hand sides of the master equation, we can now recover the simple dynamics of the Poisson distribution parameter:

$$\Delta E_{ij} = p_{ij}(t); \quad (\text{A.6})$$

$$E_{ij}(t) = \sum_{t'=1}^t p_{ij}(t'), \quad (\text{A.7})$$

where we can now call the quantity  $E_{ij}(t)$  *integral exposure* of the edge  $(i, j)$ . We showed that so long as the three assumptions hold, the distribution of memories of the edge follows the Poisson shape (Eq. A.3); the process of memory accumulation is a Poisson process. For the non-existent edges of the network  $p_{ij}(t) = 0$  for all  $t$ , and thus they never accumulate any memories.

For a dynamic network, the edge visitation probability is a function of the evolution time  $\tau$ , which is related to the random walk time via dilation  $t = D\tau$ . For long times and smooth evolution, we can approximate the sum Eq. A.7 with an integral that makes the change of variables much more straightforward:

$$E_{ij}(t) = \int_0^t p_{ij}(\tau') d\tau' = D \int_0^{t/D} p_{ij}(\tau') d\tau' = D \mathcal{E}_{ij}(t/D), \quad (\text{A.8})$$

which recovers the formula from the main text. For time-dependent networks,  $p_{ij}(\tau) = 0$  for some edges for some part of the network evolution. Thus, new memories cannot be accumulated but exposure does not decrease. Exposure thus accounts for both edge weight and time dependence.

## B. Poisson process calculus and coarse-graining

The above theory was developed assuming that the elementary events of interest are visitations of edges. However, for other applications, we might be interested in the memory of a *group* of edges  $(i, j) \in g$ . Since the visitations are independent, we can just compute the distribution of memory counts for the whole group. Poisson processes are additive, regardless of the parameter that is:

$$\sum_{ij \in g} \text{Pois}(E_{ij}(t)) = \text{Pois}\left(\sum_{ij \in g} E_{ij}(t)\right), \quad (\text{B.1})$$

where the equality states that the left and right sides of the equation have identical distributions. Drawing pseudorandom numbers from distributions on a computer is typically a computationally expensive operation (see Section H below), while addition is cheap. Drawing a realization from the left-hand side of Eq. B.1 requires doing the expensive operation once for each edge, while drawing a realization from the right-hand side requires the expensive operation only once at all. The existence of Eq. B.1 thus promises a significant quantitative and computational benefit.

The benefits of the *coarse-graining* Eq. B.1 are not only quantitative but also qualitative and conceptual as well. Since the group  $g$  can be defined arbitrarily, we can use the expression to compute the exposure  $E_g$  of different groups and attach group-specific meaning to it. The choice of group is equivalent to the choice of an order parameter in conventional Boltzmann statistical mechanics [64]. We show two particular examples of the group  $g$  below, but emphasize that other options are possible.

One particular choice of the group is all edges  $(i, j)$  that connect to a particular node  $j$ . Since traversing any of those edges is identical to visiting the node  $j$ , we can use the coarse-graining formula B.1 to find the distribution of memories of visiting a node, which provides a view of the learned network complementary to the edges. We can thus use the *edge* exposure  $E_{ij}(t)$  to compute the *node* exposure:

$$K_j(t) \equiv \sum_i E_{ij}(t) = \sum_{t'=0}^t \pi_j(t'); \quad (\text{B.2})$$

$$M_j \sim \text{Pois}(K_j(t)), \quad (\text{B.3})$$

where  $\pi_j$  is the instantaneous steady-state visitation probability of a node.

Because the node exposures are sums of non-negative edge exposures, they are typically larger than edge exposures. From the exposures, we can compute the probability of visiting a node with Eq. B.6, which would grow with time much faster than the probability of visiting edges. For a network that is either unweighted or has a narrow edge weight distribution, edge exposures are  $E_{ij} = \mathcal{O}(1/m)$ . For a network with a narrow degree distribution, node exposures are  $K_j = \mathcal{O}(1/n)$ . A random walk in such networks would visit all the nodes in  $\mathcal{O}(n \ln n)$  time and all the edges in  $\mathcal{O}(m \ln m)$  time, consistent with prior results [6, 10]. For heterogeneously structured networks, which are our focus here, the exploration is non-uniform and full exploration of the entire network can take much longer than it does in homogeneously structured networks.

Another choice of the group is to just include all edges  $(i, j)$  and thus find the total exposure and the number of memories of the random walker. By analogy with Boltzmann's statistical mechanics, we can call such a sum the partition function:

$$\mathcal{Z}(t) = \sum_{ij} E_{ij}(t) = \sum_{t'=0}^t \sum_{ij} p_{ij}(t') = t; \quad (\text{B.4})$$

$$M \sim \text{Pois}(t). \quad (\text{B.5})$$

Thus, regardless of the dynamics of network evolution the total number of memories fluctuates around the walk length  $t$ , as would be expected for a canonical ensemble.

Along with the distributions of memories of a group of edges, we can compute averages or observables over the distribution. A common quantity of interest is a binary variable of edge visitation, that is, whether there are any memories of that edge. We compute the probability of visitation from the Poisson distribution (Eq. A.3):

$$P(M_{ij} > 0, t) = 1 - P(M_{ij} = 0, t) = 1 - e^{-E_{ij}(t)}, \quad (\text{B.6})$$

which we use to compute the expected visited edge fraction and the CDFs in the two following sections, respectively.

### C. Visited edge fraction and the Jensen bound proof

Exposure theory predicts the probability of visiting any particular edge in Eq. B.6. By averaging this probability over all edges, we compute the expected fraction of all edges visited by time  $t$ :

$$V(t) = 1 - \frac{1}{m} \sum_{ij \in A_{ij}} e^{-E_{ij}(t)}, \quad (\text{C.1})$$

where the sum runs over the edges of nonzero weight. Here,  $\phi(x) = e^{-x}$  is a convex function, and thus we can apply Jensen's inequality:

$$1 - \frac{1}{m} \sum_{ij \in A_{ij}} e^{-E_{ij}(t)} \leq 1 - e^{-\frac{1}{m} \sum_{ij \in A_{ij}} E_{ij}(t)}. \quad (\text{C.2})$$

In order to simplify the expression, on the right-hand side we exchange the summation order over  $(i, j)$  and  $t$ . Since at every time step the network as a whole gets exactly 1 unit of exposure, over  $t$  time steps it gets  $\mathcal{L}(t) = t$  units of exposure, turning the bound into:

$$V(t) \leq 1 - e^{-t/m}. \quad (\text{C.3})$$

Recall that for an unweighted network, every edge has the same exposure  $E_{ij} = t/m$ . For such a network, a direct computation of the sum Eq. C.1 results in the right-hand side of Eq. C.3. In other words, random walks on unweighted networks exactly and uniquely saturate the Jensen bound of exploration.

### D. CDF computation

Here, we show how to compute the shape of the CDF shown in Fig. 3(b,c) of the main text. Since the computation proceeds identically for both edge weight  $A_{ij}$  and filtration  $F_{ij}$ , we just denote the edge variable with  $X_{ij}$ . We define the *non-decreasing ordering* of edges  $(i, j) \rightarrow q$  such that  $X_{q_1} \leq X_{q_2}, \forall (q_1 < q_2)$ . Whenever multiple edges have the same value of  $X$ , their relative order is arbitrary. Every edge has a probability  $P_k$  of being visited in a particular process. The CDF can be drawn as a parametric curve in index  $q \in [0, m]$  with the coordinates along the axes equal to:

$$\left( X_q, \sum_1^q P_{q'} / \sum_1^m P_{q'} \right). \quad (\text{D.1})$$

In order to draw this curve, we need to find the probability  $P_k$  for each of the three cases: the original textbook network, the stochastic simulation, and the exposure theory prediction. They are computed as follows:

$$P_q = \begin{cases} 1, & \text{textbook} \\ [M_q^r > 0], & \text{stochastic} , \\ (1 - e^{-E_q(t)}), & \text{exposure} \end{cases} \quad (\text{D.2})$$

where we used the *same ordering*  $(i, j) \rightarrow q$  determined from the values of  $X_{ij}$ . For the original network, every edge is present, so on the vertical axis the points of the curve Eq. D.1 are equally spaced. For

the stochastic simulation, we use the elements of the memory matrix  $M_{ij}^r$  from the run replica  $r$ . For the exposure prediction, the probability of each edge visit is given by Eqn. B.6. At very long runtimes, every edge would be visited at least once, so both the stochastic and the exposure expressions would approach 1.

### E. Exposure predicts memory counts

Another way to directly test the predictions of exposure theory is to compare the statistics of the accrued memories of each edge. The number of memories  $M_{ij}$  is always a non-negative integer. From stochastic simulations, we compute the mean and standard deviation of memories of every edge  $\langle M_{ij} \rangle, \sigma_{M_{ij}}$ . From exposure theory, the number of memories has a Poisson distribution, which has the mean and variance (first and second cumulants) equal to the parameter:

$$M_{ij} \sim \text{Pois}(E_{ij}(t)), \quad (\text{E.1})$$

$$\langle M_{ij} \rangle_c = \langle M_{ij}^2 \rangle_c = \sigma_{M_{ij}}^2 = E_{ij}(t). \quad (\text{E.2})$$

The exposure of edges varies over many orders of magnitude: some edges surely get many memory counts, while others barely get any. The threshold for edge discovery, as discussed in the main text, is  $E_{ij}(t) \simeq 1$ . If the ‘typical’ memory counts fit in the range of Poisson mean  $\pm$  standard deviation, for  $E_{ij} < 1$  this range starts including the value of 0: fluctuations in counts become larger than the mean.

We compare the memory count range between the stochastic simulations and exposure theory (Fig. E.1). For the unweighted Copperfield network (panel a), all edges have identical exposure and thus we expect all edges to have identical memory counts. We compute the memory counts at times  $t = \{m, 3m, 5m\}$ , corresponding to 1, 3 and 5 visits per edge on average. The resulting memory count distribution across  $R = 100$  replicas is indeed flat across the edge number, with a wide and uniform standard deviation range.

For the weighted US airports network, the exposure of each edge is proportional to time  $E_{ij} = t\mathcal{E}_{ij}$ , but also varies over almost five orders of magnitude (Fig. E.1(b)). Across the whole range of exposure, the stochastic memory count range follows the shape predicted by the Poisson distribution. The standard deviation range dips down to zero right at the discovery threshold  $\mathcal{E}_{\text{disc}} = 1/t$ , but for longer times  $t$  this discovery threshold moves to the left, so that edges with lower relative weight are discovered. For the US airport network, we also have a second threshold associated with the finite number of  $R = 100$  stochastic replicas. Edges with exposure of  $E = 1/R = 10^{-2}$  are expected to be seen only once in 100 replica runs; that is, they are rare events. Standard stochastic sampling that we use here has limited capacity to estimate the frequency of such rare events, but the exposure theory prediction is valid for arbitrary values of  $E$ . As the simulation runtime changes from  $t = 1 \cdot m$  to  $t = 5 \cdot m$ , the replica threshold  $\mathcal{E}_{\text{repl}} = 1/Rt$  moves to the left as well.

For the weighted and temporal Treil network, the edge exposure accumulated by the end of the textbook is proportional to the dilation  $E_{ij} = D\mathcal{E}_{ij}$  (Fig. E.1(c)). Due to longer runtimes of simulation, we only simulated  $R = 10$  replicas here. Similarly to the US airports network, the stochastic memory count range follows the shape predicted by the Poisson distribution, with more noise due to fewer replicas. Since the difference in subsequent dilation  $D$  values is a factor of 10, the discovery threshold  $\mathcal{E}_{\text{disc}} = 1/D$  shifts much more significantly between the panels than it did for the US airport network.

In conclusion, the Poisson process of memory accrual is accurate for describing not only the binary edge visitation probability but also the distribution of the number of visits. The relative fluctuations in the number of memory counts  $M$  fall off as  $1/\sqrt{M}$  for high exposure, as expected for the Poisson distribution.

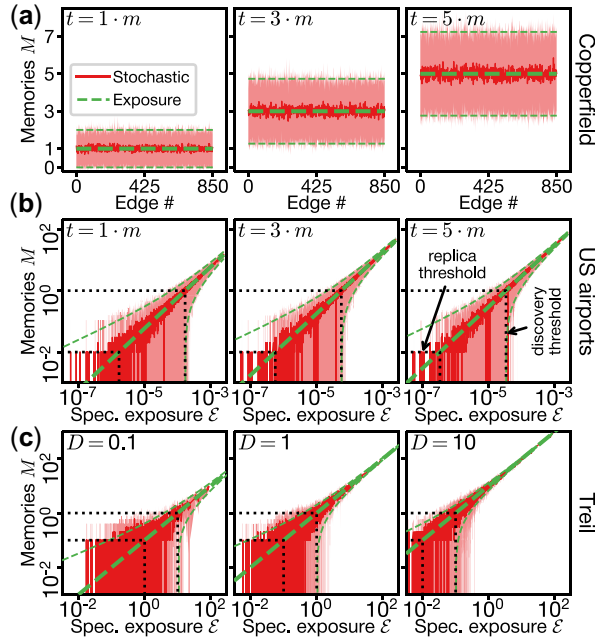


FIG. E.1. Memories accrued for every edge of the networks at different time  $t$  or dilation  $D$ . Networks include (a) David Copperfield, (b) US airports, and (c) Trell linear algebra textbook. For (a), edges are ordered arbitrarily, whereas for (b–c) edges are ordered by specific exposure  $\mathcal{E}$ . For (b,c), the horizontal and vertical dashed contours indicate the discovery threshold  $M = 1$  and the replica threshold  $M = 1/R$ . The solid line and shading are stochastic mean  $\pm$ std, respectively; the thick and thin dashed lines are the exposure mean and the  $\pm$ std range, respectively.

## F. Random walk correlation time

As a random walker explores the network, its probability of ending up on a particular node  $i$  in exactly  $t$  steps can be expressed as [1]:

$$p_i(t) = \sum_{k=0}^{n-1} a_k v_i^k \lambda_k^t, \quad (\text{F.1})$$

where  $\lambda_k$  are the eigenvalues of the transition matrix  $P(j|i)$ ,  $v_i^k$  is the  $i$ th component of the  $k$ th left eigenvector, and  $a_k$  are coefficients that depend on the initial conditions. For connected graphs, over long times the probability distribution approaches the asymptotic distribution given by the top left eigenvector  $v_i^0 = \pi_i$ . How long does that decay take?

By the Perron–Frobenius theorem, the transition matrix has the largest eigenvalue of  $\lambda_0 = 1$ , and all other eigenvalues are smaller or equal by absolute value  $|\lambda_k| \leq 1$ . The presence of multiple eigenvalues of  $\lambda_k = 1$  indicates the existence of multiple network components. All eigenvalues smaller than 1 by absolute value set up the hierarchy of timescales equal to  $t_k \simeq -1/\ln(|\lambda_k|)$  and commonly referred to as mixing, relaxation, decay or correlation times. The corresponding eigenvector indicates which nodes are involved in the relaxation mode—whether just a few or many. Typically the second largest eigenvalue  $\lambda_1$  is taken to compute the mixing, that is, correlation time. If the correlation time is close to the one step of random walk  $t_{\text{corr}} \simeq 1$ , the subsequent nodes visited are effectively independently sampled from



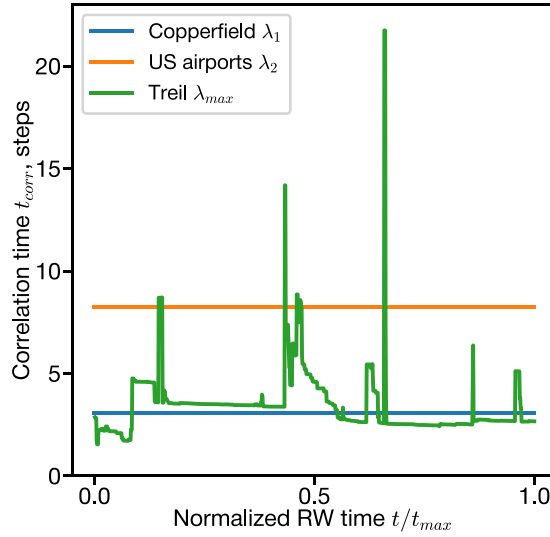


FIG. F.1. Correlation times for the three networks across the normalized random walk time. Networks include David Copperfield (bottom flat line), US airports (top flat line), and Treil linear algebra textbook (fluctuating curve) networks.

the instantaneous steady-state distribution  $\pi_i$ , and thus Assumptions 2 and 3 of exposure theory hold. However, since the three networks we study have a different nature, we re-examine the correlation time for each of them separately (Fig. F.1).

The Copperfield network is connected and unweighted. It has exactly one eigenvalue of 1, with the next eigenvalue by absolute value equal to  $\lambda_1 \simeq -0.7235$ . Note that the eigenvalue is negative, which corresponds to the underlying disassortative (nearly bipartite) structure of the graph. The network originates from the adjacency of nouns and adjectives in the text of an English novel, where each part of speech is more often paired with the opposite part; that is, there are more noun–adjective adjacencies than noun–noun and adjective–adjective adjacencies [25]. The negative eigenvalue implies that correlations decay in an oscillatory manner, but we can compute the mixing time nonetheless to be  $t_{\text{corr}} \simeq 3.09$ . Since the correlation time is close to 1 and much smaller than the typical random walk times we consider  $t_{\text{corr}} \ll t \sim m = 850$ , the assumptions of exposure theory hold for the Copperfield network.

The US airports network is also connected but has a wide range of weights. After the first eigenvalue of 1, it has the second eigenvalue of  $\lambda_1 \simeq 0.9976$ , corresponding to  $t_{\text{corr}} \simeq 419$  steps, a very large number. However, the corresponding second left eigenvector is entirely localized to 5 nodes with the lowest weighted degree (also known as the nodal *strength*). It takes over 400 steps for the random walk to discover those nodes, but they are not informative of how the rest of the network is explored. In order to estimate the speed of exploration of the rest of the network, we turn to the third eigenvalue of  $\lambda_2 \simeq 0.8857$ , corresponding to the timescale  $t_{\text{corr}} \simeq 8.24$ . The third eigenvector is delocalized across the network, corresponding to broader mixing. This correlation time is larger than for the Copperfield network, but still much smaller than the typical random walk times  $t_{\text{corr}} \ll t \sim m = 5960$ , so the assumptions of exposure theory also hold for the US airport network.

The Treil network, unlike the previous two, is temporal. As the network evolves, it changes the structure and the corresponding correlation times. We thus compute the instantaneous spectrum of the transition matrix generated by normalizing the temporal adjacency matrix  $\mathbf{A}(\tau)$ . We rescale the evolution

time to fit within  $[0, 1]$  to illustrate the dynamics. While the whole Treil network is connected, during the evolution it consists of a large connected component of most nodes and small disconnected components of a few nodes that only exist for several steps. In the transition matrix spectrum, the small components manifest through additional eigenvalues of 1 or  $-1$ . Such disconnected components are quickly connected back to the main part of the network. Therefore, in order to estimate the correlation time, we use the instantaneous eigenvalue  $\lambda_{\max}$  which is the largest by magnitude but smaller than 1. The resulting curve in Fig. F.1 mostly hovers around 3 or 4 steps, with occasional large spikes when weak connections to new nodes are introduced. The lifetime of such spikes is typically shorter than their magnitude: before the random walker finds a new node via a weak connection, the connection becomes stronger. Outside of the short-lived spikes, the correlation time stays much smaller than the typical random walk time  $t_{\text{corr}} \ll D\tau_{\max}$ , where  $\tau_{\max} = 6681$  sentences for the Treil textbook. The assumptions of exposure theory hold for the Treil network as well.

The number of memories of a particular edge can be computed as a sum of increments at each time step:

$$M_{ij}(t) = \sum_{t'=1}^t \Delta M_{ij}(t'), \quad (\text{F.2})$$

where  $\Delta M_{ij}(t')$  is a random number which is equal to 1 with low probability (when the edge  $(i, j)$  is visited) and 0 otherwise. The sequence of increments is auto-correlated on the timescale  $t_{\text{corr}}$ . We found the correlation time on all three networks to be small  $t_{\text{corr}} < 10^1$ , while the random walk simulations run for  $10^3$ – $10^4$  steps. The full sum in Eq. F.2 thus consists of many uncorrelated blocks. Moreover, in these simulations, we mainly focus on the estimator of the *mean* of accumulated memories, which is unbiased even for auto-correlated increments [65]. In conclusion, correlation time analysis shows that exposure theory assumptions are satisfied for the three studied networks at the studied random walk lengths.

## G. Breaking exposure theory

In order to clarify the domain of applicability for the exposure theory, we compare its predictions to random walks on networks that explicitly break the exposure theory assumption of small correlation time, similar to Ref. [10]. We test two groups of networks with roughly constant number of nodes. In the first group, we construct a series of five regular cubic lattices with increasing dimension and closed boundary conditions: a  $n = 256^1$  node 1D lattice, a  $n = 16^2 = 256$  node 2D lattice, a  $n = 6^3 = 216$  node 3D lattice, a  $n = 4^4 = 256$  node 4D lattice, and a  $n = 3^5 = 243$  node 5D lattice. In the second group, we construct a series of five Watts–Strogatz small-world networks of  $n = 256$  nodes with  $k = 4$  nearest neighbour connections and increasing probability of rewiring  $p \in \{0.0, 0.1, 0.2, 0.3, 0.4\}$ ; at each probability, we consider one stochastic rewiring realization in which the network remains connected [66]. On each network, we compute 100 replicas of stochastic random walk simulations, always starting from the same node. Across both groups, all networks are unweighted and undirected. Thus exposure theory has identical local and aggregate predictions in Eqs. B.6 and C.1. But do simulations follow the prediction?

While according to exposure theory all edges should have an identical learning curve, in stochastic simulations the curves differ significantly. For 1D and 2D lattices (Fig. G.1(a,b)), the closest edges have notably higher discovery probability than the exposure theory predicts, and the farthest edges have a lower probability. For the 1D network, several edges have never been discovered over 100 replicas running for time  $t = 5 \cdot m$ . In contrast, for 3D, 4D and 5D lattices, the stochastic curves closely follow the

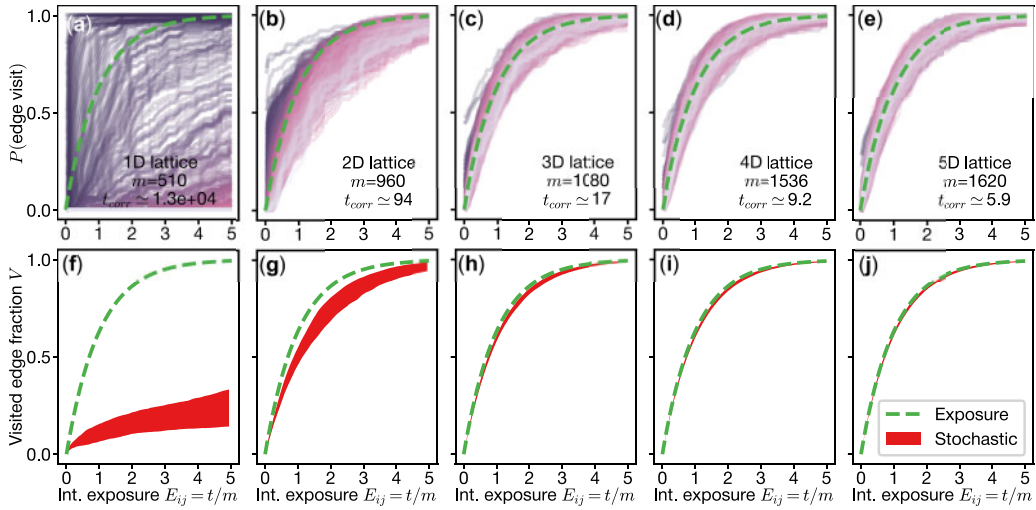


FIG. G.1. Exposure theory is progressively more accurate with increasing spatial dimension. (a–e) Probability of discovery of each network edge in lattices of increasing dimension (solid transparent curves), curve shade varies from dark for closest edges to light for farthest edges. The dashed curve corresponds to exposure prediction. (f–j) Fraction of visited edges over time for the corresponding networks. The dashed curve corresponds to exposure prediction, the shaded region corresponds to stochastic mean  $\pm$  std.

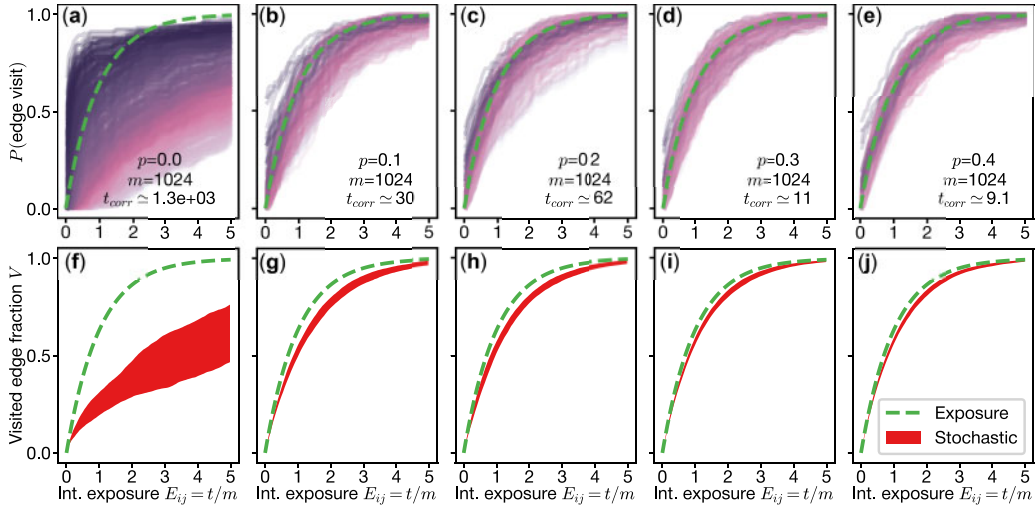


FIG. G.2. Exposure theory is progressively more accurate with increasing network randomness. (a–e) Probability of discovery of each network edge in five Watts-Strogatz networks with increasing rewiring probability  $p$  (solid transparent curves), curve shade varies from dark for closest edges to light for farthest edges. The dashed curve corresponds to exposure prediction. (f–j) Fraction of visited edges over time for the corresponding networks. The dashed curve corresponds to exposure prediction, the shaded region corresponds to stochastic mean  $\pm$  std.

exposure prediction (Fig. G.1(c–e)). We observe a similar pattern for the aggregate metric of visited edge fraction. For the 1D lattice (Fig. G.1(f)), exposure theory drastically overestimates the rate of network exploration. As the lattice dimension gets higher, the standard deviation of the stochastic exploration

curve decreases and the mean gets closer to the exposure prediction. As expected, exposure theory fails for low-dimensional lattices but works well for high-dimensional ones.

The pattern of prediction success is similar for the Watts–Strogatz networks. Before any rewiring occurs, local edge visitation curves strongly deviate from the exposure prediction and aggregate exploration is much slower than predicted (Fig. G.2(a,f)). As the network is gradually rewired, the visitation curves get more consistent and the visited edge fraction more closely follows the prediction, with smaller variance (Fig. G.2(b–e,g–j)). As expected, exposure theory fails for un-rewired Watts–Strogatz networks (effectively one-dimensional), but works well once the networks are randomized.

We thus showed a scenario in which exposure theory gives incorrect predictions. This scenario can be broken in two very different ways: either by constructing regular high-dimensional lattices, or by abandoning dimensional structure in favour of irregular rewiring. Both ways lead to reduction of the correlation time  $t_{\text{corr}}$  (Figs G.1 and G.2(a–e)). The drop in correlation time decreases the conditional dependence of edge visitations on the starting node, thus ensuring that a key assumption of exposure theory holds. Thus exploration of either high-dimensional lattices or random networks is qualitatively and quantitatively similar to the exploration of the complex networks shown in the main text.

## H. Computational benchmark

In previous sections, we established the accuracy of exposure theory predictions. However, how does exposure theory compare to stochastic simulations in terms of computational resources?

All three networks we consider in this article are sparse; that is, the number of edges is significantly smaller than the number of possible node pairs  $m \ll n^2$ . Because of this, the adjacency matrix  $\mathbf{A}$ , the filtration matrix  $\mathbf{F}$ , the memory matrix  $\mathbf{M}$ , and the exposure matrix  $E$  are all sparse and can be stored in  $\mathcal{O}(m)$  memory space.

The time complexity of a stochastic random walk simulation is linear in the random walk time  $\mathcal{O}(t)$ . However, as we established above, the relevant timescale for random walks is about the number of edges  $m$ . Let's set  $t = h \cdot m$ , where  $h$  is a small number ( $1 \leq h \leq 5$ ). The most computationally complex part of the random walk simulation is generating pseudorandom numbers to decide on the next edge to traverse, one per step. In order to save on this expense, we run the random walk simulation once for every replica and save the whole trajectory (the sequence of visited nodes). We then perform a variety of data analyses on the random walk reconstructed from the saved trajectory. While both the pseudorandom and the reconstructed trajectories give the same deterministic result (the memory matrix  $\mathbf{M}$ ), the latter is drastically faster.

The exposure theory requires computing the exposure value for each edge. For static networks, computing the specific exposure  $\mathcal{E}_{ij}$  requires dividing the weight of each edge by the sum of all weights and thus takes  $\mathcal{O}(m)$  time. For time-dependent networks, computing the specific exposure  $\mathcal{E}_{ij}(\tau)$  requires evaluating the integral in Eq. A.8. For slowly varying networks, it can be computed with a simple rectangle rule and thus has time complexity  $\mathcal{O}(m \cdot n_{st})$ , where  $n_{st}$  is the number of computational steps in time integration which we set to  $n_{st} = 10^3$ . Once the specific exposure  $\mathcal{E}_{ij}$  is known, it can be converted into integral exposure  $E_{ij}$  through multiplying it by a scalar random walk time  $t$  or dilation  $D$ . Finding the visitation probability via Eq. B.6 or the fraction of node visits via Eq. C.1 reduces to standard numerical algebra, which is optimized in modern computing packages.

We have thus established that stochastic simulation, reconstruction, and exposure computation all scale as  $\mathcal{O}(m)$ . We then run a benchmark of a Python implementation of the algorithms on a laptop computer (Intel Core i7-1065G7 @ 1.30GHz, 16Gb RAM) to get the absolute values of time, presented in Fig. H.1.

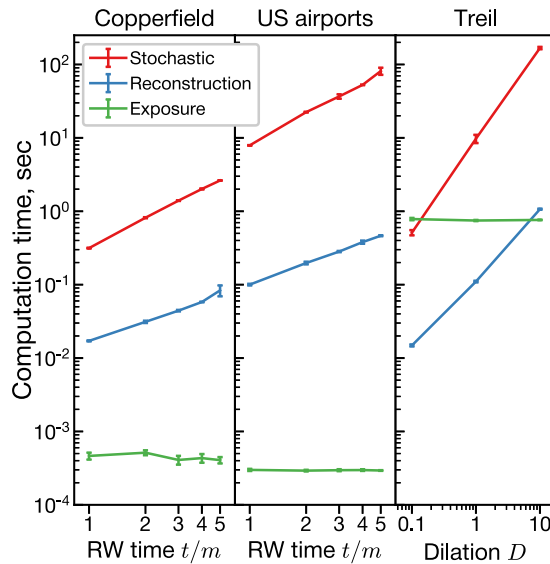


FIG. H.1. Computation time of the three algorithms on each of the three networks. Error bars indicate standard deviation over 5 runs. Note that the horizontal axis is logarithmic and spans the range  $[1, 5]$  for the first two networks, but a much larger range  $[0.1, 10]$  for the third network.

Both stochastic simulation and reconstruction scale linearly with either random walk time  $h = t/m$  or dilation  $D$ , but reconstruction is faster than stochastic simulation by a factor of  $10^1$ – $10^2$ . In contrast, the exposure computation runtime does not scale with either random walk time or dilation and instead takes constant time. For static networks, using exposure theory is faster than computing *one replica* of stochastic simulation by a factor of  $10^3$ – $10^5$ , depending on the desired random walk time and network size. For the dynamic Treil network, using  $n_{st} = 10^3$  integration steps slows down the algorithm by the corresponding factor. However, the specific exposure computation can be performed once and stored in a file for fast lookup.

The specific runtime of exposure theory computations depends on the fine details of implementation, programming language choice, numerical linear algebra, memory calls and data structures, and other low-level optimization. The green curve on Fig. H.1 thus represents not the ultimate bound of possible performance, but merely the speed achieved by the authors of the present study. The largest gain in performance comes from not drawing pseudorandom realizations of random walk steps, but treating the probabilities as floating-point numbers with fast algebra. Working directly with probability values also obviated the need to collect many stochastic random walk samples. Since getting the statistics to validate exposure predictions in this paper required between  $10^1$  and  $10^2$  independent samples, adopting exposure theory can yield a speedup by a factor of a million.

#### REFERENCES

1. MASUDA, N., PORTER, M. A. & LAMBIOTTE, R. (2017) Random walks and diffusion on networks. *Phys. Rep.*, **716**, 1–58.

2. DVORETZKY, A. & ERDÖS, P. (1951) Some problems on random walk in space. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (J. Neyman ed.). Berkeley and Los Angeles, California: University of California Press, pp. 353–367.
3. PÓLYA, G. (1921) Über eine aufgabe der wahrscheinlichkeitsrechnung betreffend die irrfahrt im straßennetz. *Math. Ann.*, **84**, 149–160.
4. DA FONTOURA COSTA, L. (2006) Learning about knowledge: a complex network approach. *Phys. Rev. E*, **74**, 026103.
5. DA FONTOURA COSTA, L. & TRAVIESO, G. (2007) Exploring complex networks through random walks. *Phys. Rev. E*, **75**, 016102.
6. ASZTALOS, A. & TOROCZKAI, Z. (2010) Network discovery by generalized random walks. *Europhys. Lett.*, **92**, 50008.
7. VOLCHENKOV, D. (2011) Random walks and flights over connected graphs and complex networks. *Commun. Nonlinear Sci. Numer. Simul.*, **16**, 21–55.
8. COOPER, C. & FRIEZE, A. (2007) The cover time of sparse random graphs. *Random Struct. Algorithms*, **30**, 1–16.
9. COOPER, C. & FRIEZE, A. (2007) The cover time of the preferential attachment graph. *J. Combin. Theory, Ser. B*, **97**, 269–290.
10. MAIER, B. F. & DIRK BROCKMANN. (2017) Cover time for random walks on arbitrary complex networks. *Phys. Rev. E*, **96**, 042307.
11. DELVENNE, J.-C. & LIBERT, A.-S. (2011) Centrality measures and thermodynamic formalism for complex networks. *Phys. Rev. E*, **83**, 046117.
12. ROSVALL, M. & BERGSTROM, C. T. (2008) Maps of random walks on complex networks reveal community structure. *Proc. Nat. Acad. Sci. USA*, **105**, 1118–1123.
13. PICCARDI, C. (2011) Finding and testing network communities by lumped Markov chains. *PLoS One*, **6**, e27028.
14. SINATRA, R., GÓMEZ-GARDENES, J., LAMBIOTTE, R., NICOSIA, V. & LATORA, V. (2011) Maximal-entropy random walks in complex networks with limited information. *Phys. Rev. E*, **83**, 030103.
15. DE ARRUDA, H. F., SILVA, F. N., DA F. COSTA, L. & AMANCIO, D. R. (2017) Knowledge acquisition: a complex networks approach. *Inf. Sci.*, **421**, 154–166.
16. IACOPINI, I., MILOJEVIĆ, S. & LATORA, V. (2018) Network dynamics of innovation processes. *Phys. Rev. Lett.*, **120**, 048301.
17. SCHAPIRO, A. C., ROGERS, T. T., CORDOVA, N. I., TURK-BROWNE, N. B. & BOTVINICK, M. M. (2013) Neural representations of events arise from temporal community structure. *Nat. Neurosci.*, **16**, 486–492.
18. STACHENFELD, K. L., BOTVINICK, M. M. & GERSHMAN, S. J. (2017) The hippocampus as a predictive map. *Nat. Neurosci.*, **20**, 1643–1653.
19. LYNN, C. W. & BASSETT, D. S. (2020) How humans learn and represent networks. *Proc. Natl. Acad. Sci. USA*, **117**, 29407–29415.
20. KARUZA, E. A. (2022) The value of statistical learning to cognitive network science. *Top. Cogn. Sci.*, **14**, 78–92.
21. LYNN, C. W., KAHN, A. E., NYEMA, N. & BASSETT, D. S. (2020) Abstract representations of events arise from mental errors in learning and memory. *Nat. Commun.*, **11**, 1–12.
22. LYNN, C. W., PAPADOPOULOS, L., KAHN, A. E. & BASSETT, D. S. (2020) Human information processing in complex networks. *Nat. Phys.*, **16**, 965–973.
23. ALMAAS, E., KULKARNI, R. V. & STROUD, D. (2003) Scaling properties of random walks on small-world networks. *Phys. Rev. E*, **68**, 056105.
24. NEWMAN, M. E. J. (2006) Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA*, **103**, 8577–8582.
25. NEWMAN, M. E. J. (2006) Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E*, **74**, 036104.
26. COLIZZA, V., PASTOR-SATORRAS, R. & VESPIGNANI, A. (2007) Reaction–diffusion processes and metapopulation models in heterogeneous networks. *Nat. Phys.*, **3**, 276–282.
27. CHRISTIANSON, N. H., BLEVINS, A. S. & BASSETT, D. S. (2020) Architecture and evolution of semantic networks in mathematics texts. *Proc. R. Soc. A*, **476**, 20190741.



28. LAU, H. W. & SZETO, K. Y. (2010) Asymptotic analysis of first passage time in complex networks. *Europhys. Lett.*, **90**, 40005.
29. BARTOLUCCI, S., CACCIOLI, F., CARAVELLI, F. & VIVO, P. (2021) “Spectrally gapped” random walks on networks: a mean first passage time formula. *SciPost Phys.*, **11**, 088.
30. CLAUSET, A. & MOORE, C. (2003) Traceroute sampling makes random graphs appear to have power law degree distributions. *arXiv preprint cond-mat/0312674*.
31. ACHLIOPTAS, D., CLAUSET, A., KEMPE, D. & MOORE, C. (2009) On the bias of traceroute sampling: or, power-law degree distributions in regular graphs. *J. ACM*, **56**, 1–28.
32. STUTZBACH, D., REJAIE, R., DUFFIELD, N., SEN, S. & WILLINGER, W. (2008) On unbiased sampling for unstructured peer-to-peer networks. *IEEE/ACM Trans. Netw.*, **17**, 377–390.
33. LAMBIOTTE, R., DELVENNE, J.-C. & BARAHONA, M. (2014) Random walks, Markov processes and the multiscale modular organization of complex networks. *IEEE Trans. Netw. Sci. Eng.*, **1**, 76–90.
34. SCHAUB, M. T., DELVENNE, J.-C., YALIRAKI, S. N. & BARAHONA, M. (2012) Markov dynamics as a zooming lens for multiscale community detection: non clique-like communities and the field-of-view limit. *PLoS One*, **7**, e32210.
35. ARNAUDON, A., PEACH, R. L. & BARAHONA, M. (2020) Scale-dependent measure of network centrality from diffusion dynamics. *Phys. Rev. Res.*, **2**, 033104.
36. GHAVASIEH, A., NICOLINI, C. & DE DOMENICO, M. (2020) Statistical physics of complex information dynamics. *Phys. Rev. E*, **102**, 052304.
37. GHAVASIEH, A. & DE DOMENICO, M. (2022) Statistical physics of network structure and information dynamics. *J. Phys.*, **3**, 011001.
38. RAHMEDE, C., IACOVACCI, J., ARENAS, A. & BIANCONI, G. (2018) Centralities of nodes and influences of layers in large multiplex networks. *J. Complex Netw.*, **6**, 733–752.
39. BIANCONI, G. & RAHMEDE, C. (2017) Emergent hyperbolic network geometry. *Sci. Rep.*, **7**, 1–9.
40. MILLÁN, A. P., TORRES, J. J. & BIANCONI, G. (2019) Synchronization in network geometries with finite spectral dimension. *Phys. Rev. E*, **99**, 022307.
41. MILLÁN, A. P., GHORBANCHIAN, R., DEFENU, N., BATTISTON, F. & BIANCONI, G. (2021) Local topological moves determine global diffusion properties of hyperbolic higher-order networks. *Phys. Rev. E*, **104**, 054302.
42. HUFNAGEL, L., BROCKMANN, D. & GEISEL, T. (2004) Forecast and control of epidemics in a globalized world. *Proc. Natl. Acad. Sci. USA*, **101**, 15124–15129.
43. BROCKMANN, D. & HELBING, D. (2013) The hidden geometry of complex, network-driven contagion phenomena. *Science*, **342**, 1337–1342.
44. IANNELLI, F., KOHER, A., BROCKMANN, D., HÖVEL, P. & SOKOLOV, I. M. (2017) Effective distances for epidemics spreading on complex networks. *Phys. Rev. E*, **95**, 012313.
45. BONAVENTURA, M., NICOSIA, V. & LATORA, V. (2014) Characteristic times of biased random walks on complex networks. *Phys. Rev. E*, **89**, 012803.
46. HOLME, P. & SARAMÄKI, J. (2012) Temporal networks. *Phys. Rep.*, **519**, 97–125.
47. PERRA, N., BARONCHELLI, A., MOCANU, D., GONÇALVES, B., PASTOR-SATORRAS, R. & VESPIGNANI, A. (2012) Random walks and search in time-varying networks. *Phys. Rev. Lett.*, **109**, 238701.
48. PETIT, J., LAMBIOTTE, R. & CARLETTI, T. (2019) Classes of random walks on temporal networks with competing timescales. *Appl. Netw. Sci.*, **4**, 1–20.
49. ADORF, C. S., DODD, P. M., RAMASUBRAMANI, V. & GLOTZER, S. C. (2018) Simple data and workflow management with the signac framework. *Comput. Mater. Sci.*, **146**, 220–229.
50. ADORF, C. S., RAMASUBRAMANI, V., DICE, B. D., HENRY, M. M., DODD, P. M. & GLOTZER, S. C. (2019) glotzerlab/signac.
51. MITCHELL, S. M., LANGE, S. & BRUS, H. (2013) Gendered citation patterns in international relations journals. *Int. Stud. Perspect.*, **14**, 485–492.
52. DION, M. L., SUMNER, J. L. & MITCHELL, S. M. (2018) Gendered citation patterns across political science and social science methodology fields. *Polit. Anal.*, **26**, 312–327.

53. CAPLAR, N., TACCHELLA, S. & BIRRER, S. (2017) Quantitative evaluation of gender bias in astronomical publications from citation counts. *Nat. Astronomy*, **1**, 0141.
54. MALINIAK, D., POWERS, R. & WALTER, B. F. (2013) The gender citation gap in international relations. *Int. Org.*, **67**, 889–922.
55. DWORKIN, J. D., LINN, K. A., TEICH, E. G., ZURN, P., SHIOHARA, R. T. & BASSETT, D. S. (2020) The extent and drivers of gender imbalance in neuroscience reference lists. *Nat. Neurosci.*, **23**, 918–926.
56. BERTOLERO, M. A., DWORKIN, J. D., DAVID, S. U., LLOREDA, C. L., SRIVASTAVA, P., STISO, J., ZHOU, D., DZIRASA, K., FAIR, D. A., KACZKURKIN, A. N., MARLIN, B. J., SHOHAMY, D., UDDIN, L. Q., ZURN, P. & BASSETT, D. S. (2020) Racial and ethnic imbalance in neuroscience reference lists and intersections with gender. *bioRxiv*.
57. WANG, X., DWORKIN, J. D., ZHOU, D., STISO, J., FALK, E. B., BASSETT, D. S., ZURN, P. & LYDON-STALEY, D. M. (2021) Gendered citation practices in the field of communication. *Ann. Int. Commun. Assoc.*, **45**, pp. 134–153.
58. CHATTERJEE, P. & WERNER, R. M. (2021) Gender disparity in citations in high-impact journal articles. *JAMA Netw. Open*, **4**, e2114509.
59. FULVIO, J. M., AKINNOLA, I. & POSTLE, B. R. (2021) Gender (im)balance in citation practices in cognitive neuroscience. *J. Cogn. Neurosci.*, **33**, 3–7.
60. TEICH, E. G., KIM, J. Z., LYNN, C. W., SIMON, S. C., KLISHIN, A. A., SZYMULA, K. P., SRIVASTAVA, P., BASSETT, L. C., ZURN, P., DWORKIN, J. D. & BASSETT, D. S. (2021) Citation inequity and gendered citation practices in contemporary physics. *arXiv preprint arXiv:2112.09047*.
61. ZHOU, D., CORNBLATH, E. J., STISO, J., TEICH, E. G., DWORKIN, J. D., BLEVINS, A. S. & BASSETT, D. S. (2020) Gender diversity statement and code notebook v1.0. <https://doi.org/10.5281/zenodo.3672110>.
62. AMBEKAR, A., WARD, C., MOHAMMED, J., MALE, S. & SKIENA, S. (2009) Name-ethnicity classification from open sources. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, US: Association for Computing Machinery, pp. 49–58.
63. SOOD, G. & LAOHAPRAPANON, S. (2018) Predicting race and ethnicity from the sequence of characters in a name. *arXiv preprint arXiv:1805.02109*.
64. GOLDENFELD, N. (1992) *Lectures on Phase Transitions and the Renormalization Group*. Reading MA: Addison-Wesley.
65. NEWMAN, M. E. J. & BARKEMA, G. T. (1999) *Monte Carlo Methods in Statistical Physics*. Oxford, UK and New York, NY: Clarendon Press.
66. WATTS, D. J. & STROGATZ, S. H. (1998) Collective dynamics of “small-world” networks. *Nature*, **393**, 440–442.