

A random walk perspective on hide-and-seek games

Shubham Pandey¹ and Reimer Kühn¹

Mathematics Department, King's College London, Strand, London WC2R 2LS, United Kingdom

E-mail: reimer.kuehn@kcl.ac.uk

Received 25 September 2018, revised 21 December 2018

Accepted for publication 8 January 2019

Published 28 January 2019



CrossMark

Abstract

We investigate hide-and-seek games on complex networks using a random walk framework. Specifically, we investigate the efficiency of various degree-biased random walk search strategies to locate items that are randomly hidden on a subset of vertices of a random graph. Vertices at which items are hidden in the network are chosen at random as well, though with probabilities that may depend on degree. We pitch various hide and seek strategies against each other, and determine the efficiency of search strategies by computing the average number of hidden items that a searcher will uncover in a random walk of n steps. Our analysis is based on the cavity method for finite single instances of the problem, and generalises previous work of De Bacco *et al* (2015 *J. Phys. A: Math. Theor.* **48** 205004) so as to cover degree-biased random walks. We also extend the analysis to deal with the thermodynamic limit of infinite system size. We study a broad spectrum of functional forms for the degree bias of both the hiding and the search strategy and investigate the efficiency of families of search strategies for cases where their functional form is either matched or unmatched to that of the hiding strategy. Our results are in excellent agreement with those of numerical simulations. We propose two simple approximations for predicting efficient search strategies. One is based on an equilibrium analysis of the random walk search strategy. While not exact, it produces correct orders of magnitude for parameters characterising optimal search strategies. The second exploits the existence of an effective drift in random walks on networks, and is expected to be efficient in systems with low concentration of small degree nodes.

Keywords: random walks, complex networks, search, network exploration, random matrices

(Some figures may appear in colour only in the online journal)

¹ Author to whom any correspondence should be addressed.

1. Introduction

There are many real-life scenarios where one party attempts to hide information that is desired by another party. Examples include hiding confidential information at a node in a computer cluster, hiding an item at a physical site, and trying to keep an object safe from a potential attack on a group of sites. The party hiding the information could have good or bad intention. E.g. they could be storing personal information, and the searcher could be a hacker trying to locate and exploit it; conversely, the hider could be stashing away stolen goods, and the seeker a police force looking for it.

The above scenarios have been formalised by an agent-based game called hide-and-seek. In an abstract formulation, there are two agents, a hider and a searcher, and a sample space formalised as a network. The hider conceals a given set of items on a subset of vertices of the network. The searcher then tries to locate those objects by searching the network.

Hide-and-seek games are a form of search games [1, 2]—a broad term that is used to describe games that involve an agent searching for something in a sample space. A multitude of search games have been explored over the years and they find applications in many fields [1]. Search games can be applied for monitoring patrolling situations [3], controlling urban security [4], controlling contagion [5], and detecting malicious packets in computer networks [6]. Chapman *et al* [7] study hide-and-seek games to address issues in cyber security, and we draw inspiration from their work.

In the present paper we analyse hide-and-seek games from a random walk perspective. I.e. we shall not be specifically concerned with game theoretic aspects, such as existence or multiplicity of Nash equilibria. Rather we propose to analyse the *efficiencies* of families of random search strategies, formalised as degree-biased random walks, when applied to locate a set of items hidden in a network according to a probabilistic hiding strategy. We take the random hiding strategies to be degree-biased as well.

The problem of a random walker exploring a network has found applications in many fields [8], including diffusion [9], infection dynamics in social networks [10, 11], or in image segmentation [12].

Random walks have been studied extensively over the years. Indeed, as emphasised by Lovász [13], there is ‘not much difference between the theory of random walks on graphs and the theory of finite Markov chains’, and so it will not come as a surprise that properties of random walks in complex networks have been studied in their own right [13–18].

In order to analyse the efficiency of a random search strategy, we adapt a result of De Bacco *et al* [19] in which the average number of different vertices of a complex network visited by a random walker performing an unbiased n -step random walk is computed. We generalise their work by considering more general degree-biased transition probabilities and use this to assess the efficiency of a family of degree-biased random search algorithms to locate items that are randomly hidden on a subset of vertices of a random graph. We also extend the finite single instance analysis of [19] to cover the thermodynamic limit of infinite system size, using the method of [20] to isolate contributions of the giant component of the system.

We compute, analyse and compare search efficiencies across a broad spectrum of search and hiding strategies. We explore a few basic types of graphs in the configuration model class, namely Erdős–Rényi graphs, random regular graphs, and scale-free graphs. The principal reason for including the latter in our analysis is that many technical, social and biological real-life networks are indeed thought to be scale-free [21]. We compare our results with those obtained using random walk simulations on large finite instances and find them in excellent agreement with those obtained using the theoretical tools developed in the present paper.

The remainder of our paper is organised as follows. In section 2, we present the random walk framework in terms of which we are going to analyse the efficiency of random search strategies, as well as the families of degree-biased strategies for hiding and searching we will investigate. In section 3 we describe the method that we use for computing the efficiency of a searcher, both for large single instances (section 3.1) and in the thermodynamic limit of infinite system size (section 3.2). In section 4, we present and discuss the results obtained, and close with a summary and concluding remarks in section 5.

2. Random walk framework

2.1. Graphs, random search strategies and their analysis

We will investigate the efficiency of search strategies using random graphs as search spaces. A random graph G is defined by a set \mathcal{V} of vertices and a set \mathcal{E} of edges represented by an adjacency matrix $C = (c_{ij})$, with its entries c_{ij} randomly taking the value 1 if nodes i and j are connected by an edge, and 0 otherwise. We denote by $N = |\mathcal{V}|$ the number of vertices or nodes in the graph. We assume the networks to be undirected, so $c_{ij} = c_{ji}$ for each pair of nodes (i, j) , and that there are no self-loops in the system, hence $c_{ii} = 0$ for all i .

Our analysis of the efficiency of random search strategies will be based on recent work of De Bacco *et al* [19] who analyse the average number $S_i(n)$ of *different* sites visited by a random walker starting on vertex i in a random walk of n steps, when n becomes large. They express $S_i(n)$ as

$$S_i(n) = \sum_{j \in \mathcal{V}} H_{ij}(n), \quad (1)$$

where $H_{ij}(n)$ denotes the probability of visiting node j at least once in the first n time steps when the walker started at node i . They evaluate the large n asymptotics of this number in terms of its z -transform

$$\hat{S}_i(z) = \sum_{n=0}^{\infty} S_i(n)z^n = \sum_{j \in \mathcal{V}} \hat{H}_{ij}(z). \quad (2)$$

The z -transform $\hat{H}_{ij}(z)$ in turn is expressed in terms of the z -transform of the n -step transition probability $G_{ij}(n) = (W^n)_{ij}$, with $W = (W_{ij})$ denoting the matrix of probabilities for one-step transitions $i \rightarrow j$. De Bacco *et al* [19] find

$$\hat{H}_{ij}(z) = \frac{1}{1-z} \frac{\hat{G}_{ij}(z)}{\hat{G}_{jj}(z)}. \quad (3)$$

In order to make this paper reasonably self-contained, we reproduce the key steps of this derivation in appendix A. The large n asymptotics of $S_i(n)$ is then extracted by analysing the $z \rightarrow 1$ asymptotics of its z -transform. The analysis in [19] covers the case where the random walker is performing an unbiased random walk, for which the probability of transitioning from node i to node j is given by

$$W_{ij} = \frac{c_{ij}}{k_i}, \quad (4)$$

where $k_i = |\partial i|$ is the degree of node i , with ∂i denoting the set of the neighbours of node i .

Our analysis of the efficiency of random search strategies can fully utilise this theoretical framework. It only requires two extensions.

The first is minor. We mark a subset of vertices of the graph as having items of interest hidden on them. To do so, we associate an indicator variable ξ_j with each site j which designates whether an item is hidden on that node ($\xi_j = 1$), or not ($\xi_j = 0$). Then

$$S_i(\boldsymbol{\xi}, n) = \sum_{j \in \mathcal{V}} H_{ij}(n) \xi_j, \quad (5)$$

with $\boldsymbol{\xi} = \{\xi_1, \xi_2, \dots, \xi_N\}$, will denote the average number of hidden items found in an n -step random walk starting at node i . Its large- n asymptotics will again be analysed in terms of the $z \rightarrow 1$ -asymptotics of its z -transform, which—using equations (2) and (3)—we see to be given by

$$\hat{S}_i(\boldsymbol{\xi}, z) = \frac{1}{1-z} \sum_{j \in \mathcal{V}} \frac{\hat{G}_{ij}(z)}{\hat{G}_{jj}(z)} \xi_j. \quad (6)$$

The second modification is concerned with looking at a wider family of random walk models. Rather than restricting the analysis to unbiased random walks, we will look at a large family of degree-biased random walks with one step transition matrices given by

$$W_{ij} = \frac{c_{ij}s(k_j)}{\Gamma_i}. \quad (7)$$

Here $s(k)$ is a function of the degree, which we will refer to as a *search strategy*. The constant Γ_i is dictated by the normalisation requirement for transition probabilities, giving

$$\Gamma_i = \sum_{j \in \mathcal{V}} c_{ij}s(k_j). \quad (8)$$

In order to evaluate (6), we need the z -transform of the matrix of n -step transition probabilities. It is given by

$$\hat{G}(z) = [\mathbb{I} - zW]^{-1}, \quad (9)$$

with \mathbb{I} denoting the $N \times N$ identity matrix. We will refer to $\hat{G}(z)$ as the resolvent of W . The matrix W of one-step transition probabilities satisfies a detailed balance condition with the equilibrium distribution

$$\pi_i = \frac{1}{Y} \Gamma_i s(k_i), \quad \text{with} \quad Y = \sum_{j \in \mathcal{V}} s(k_j) \Gamma_j. \quad (10)$$

This can be used to express $\hat{G}(z)$ in terms of a symmetric matrix $\hat{R}(z)$ as

$$\hat{G}(z) = D^{-\frac{1}{2}} \hat{R}(z) D^{\frac{1}{2}}, \quad (11)$$

where $D = \text{diag}(\Gamma_i s(k_i))$, and

$$\hat{R}(z) = [\mathbb{I} - zD^{\frac{1}{2}}WD^{-\frac{1}{2}}]^{-1}. \quad (12)$$

This matrix, the resolvent of $\mathcal{W} = D^{\frac{1}{2}}WD^{-\frac{1}{2}}$, is easily seen to be symmetric. The fact that $\hat{G}(z)$ and $\hat{R}(z)$ are related by a similarity transformation can be exploited [19] to analyse $\hat{G}(z)$ via the spectral decomposition of $\hat{R}(z)$. One has

$$\hat{R}(z) = \frac{\mathbf{v}_1 \mathbf{v}_1^T}{1-z} + \sum_{\nu=2}^N \frac{\mathbf{v}_\nu \mathbf{v}_\nu^T}{1-z\lambda_\nu} \equiv \frac{\mathbf{v}_1 \mathbf{v}_1^T}{1-z} + \hat{C}(z), \quad (13)$$

where the λ_ν and the \mathbf{v}_ν are, respectively, the eigenvalues and eigenvectors of \mathcal{W} . We have isolated the contribution of the Perron–Frobenius eigenvalue $\lambda_1 = 1$ of \mathcal{W} (hence of W), and introduced $\hat{C}(z)$ to denote the contributions corresponding to the remaining eigenvalues. The (normalised) Perron–Frobenius eigenvector \mathbf{v}_1 of \mathcal{W} has entries

$$v_{1,i} = \sqrt{\pi_i} = \sqrt{\frac{s(k_i)\Gamma_i}{Y}}. \quad (14)$$

Assuming that the graph G is connected, we know that the Markov chain described by W is irreducible, hence that the multiplicity of the largest eigenvalue is 1 and that $\lambda_\nu < 1$ for all $\nu \neq 1$ by the Perron–Frobenius theorem. In the $z \rightarrow 1$ limit, the contribution from the second term of the RHS of equation (13) is therefore negligible in comparison to the contribution from the first term.

Following the reasoning of De Bacco *et al* [19], one can use this fact to determine the dominant $z \rightarrow 1$ asymptotics of $\hat{S}_i(\boldsymbol{\xi}, z)$ in the large N limit (recall that $|\mathcal{V}| = N$), and finds

$$\hat{S}_i(\boldsymbol{\xi}, z) \sim \frac{1}{(1-z)^2 Y} \sum_{j \in \mathcal{V}} \frac{s(k_j)\Gamma_j}{\hat{R}_{jj}} \xi_j, \quad z \rightarrow 1. \quad (15)$$

In equation (15), the $N \rightarrow \infty$ -limit is assumed to be taken, and we have introduced

$$\hat{R}_{jj} = \lim_{z \rightarrow 1} \lim_{N \rightarrow \infty} \hat{R}_{jj}(z). \quad (16)$$

For the sake of completeness, the key steps of this derivation are reproduced in appendix B.

Upon taking an inverse z -transform, the $1/(1-z)^2$ divergence in equation (15) translates into a linear large- n behaviour of $S_i(\boldsymbol{\xi}, n)$ of the form

$$S_i(\boldsymbol{\xi}, n) \sim B n, \quad n \gg 1, \quad (17)$$

with

$$B = \frac{1}{Y} \sum_{j \in \mathcal{V}} \frac{s(k_j)\Gamma_j}{\hat{R}_{jj}} \xi_j. \quad (18)$$

Once more it is assumed that the $N \rightarrow \infty$ -limit is taken in this expression. Note that $S_i(\boldsymbol{\xi}, n)$ is for large n independent of the starting vertex i . We will in what follows refer to the constant B as the *search efficiency*.

The non-trivial element in the evaluation of the search efficiency B is related to the \hat{R}_{jj} that appear in the result, which according to equation (12) are the diagonal elements of the inverse of a large matrix. We adopt the approach of [19] to evaluate these diagonal elements of inverse matrices in terms of single-site variances of a suitable multivariate Gaussian distribution, and use the cavity method to do this in practice for large systems. The method will be explained below in section 3. Before that, though, we turn to describing the hiding strategies that we consider in the present paper.

2.2. Hiding strategies

In order to be able to discuss the efficiency of search strategies, we also need to specify the strategies according to which items are hidden in a network. We shall take these hiding strategies to be probabilistic as well. One of the simplest choices is unbiased random hiding, which can be characterised in terms of a Bernoulli distribution for the ξ_j as

Table 1. Overview of the degree-biased hiding and search strategies and their parametrisations investigated in the present paper.

Functional form	Hiding	Searching
Power-law	$h(k) = k^\beta$	$s(k) = k^\alpha$
Exponential	$h(k) = e^{\beta k}$	$s(k) = e^{\alpha k}$
Logarithmic	$h(k) = \log(1 + \beta k^{\gamma_h})$	$s(k) = \log(1 + \alpha k^{\gamma_s})$

$$p(\xi_j) = \rho_h \delta_{\xi_j,1} + (1 - \rho_h) \delta_{\xi_j,0} \quad (19)$$

for $0 < \rho_h \leq 1$. Here $\delta_{a,b}$ is the Kronecker–delta, i.e. $\delta_{a,b} = 1$ if $a = b$, and $\delta_{a,b} = 0$ if $a \neq b$. The parameter ρ_h specifies the average fraction of vertices which have items hidden on them. As for search, we will look at a broader family of hiding strategies that are taken to be correlated with degree, i.e. we will choose

$$p(\xi_j = 1 | k_j = k) = \rho_h \frac{h(k)}{\langle h \rangle}, \quad (20)$$

in which $h(k)$ is a function of the degree, while $\langle h \rangle$ denotes the average of $h(k)$ over the degree distribution p_k , that is $\langle h \rangle = \sum_k p_k h(k)$. In what follows, we will refer to the function $h(k)$ as a *hiding strategy*. Note that for a given hiding strategy the range of achievable ρ_h is bounded, as one needs to ensure that

$$\max_k \rho_h \frac{h(k)}{\langle h \rangle} \leq 1. \quad (21)$$

In table 1 we list the families of degree biased hiding and search strategies that we will consider in the present paper, along with the parametrisations we use to explore each family. We have tried to cover the major functional forms for both hiding and search strategies.

We will investigate the effect of selecting a particular parameterised family of search strategies for any of the given hiding strategies and look at the dependence of the search efficiency B on the parameters characterising the search strategy. We will explore both matched and mismatched combinations of hiding and search strategies. Results will be presented in section 4.

3. Evaluation of search efficiencies

In this section we turn to the actual evaluation of search efficiencies. As mentioned above, the non-trivial problem that must be solved in this evaluation is that of evaluating diagonal elements of the resolvent, i.e. diagonal element of an inverse of a large matrix.

We will perform our analysis both for single large instances of the problem by suitably adapting the cavity method developed for spectra of sparse symmetric random matrices [22], and in the thermodynamic limit by interpreting the self-consistency equations for the inverse cavity variances arising from the cavity analysis as stochastic recursions.

The single instance calculation, too, will very closely follow [19], implementing the two extensions required to handle the search aspect and the more general (degree biased) random walk models we are looking at in the present paper. The transition to the thermodynamic limit turns out to be non-trivial, though, and requires the introduction of an infinite family of (degree dependent) distributions of inverse cavity variances.

Another theoretical problem that arises is related to the fact that, whereas the single-instance cavity analysis is performed on a simply connected network (in the case of random graph ensembles, on their giant component), the analysis of the thermodynamic limit, if naively performed, includes contributions from finite clusters of the system, which would have to give null-contributions to overall search efficiencies. We will utilise methods recently devised in [20], to obtain results for the thermodynamic limit which are properly restricted to the giant component.

3.1. Cavity method

The evaluation of the \hat{R}_{ij} appearing in equation (15) requires a matrix inversion (12), followed by taking suitable limits (16). The matrix inversion would be computationally expensive for large systems. It has been pointed out in the context of evaluating spectra of random matrices [23] that expressing elements of inverse matrices as covariances of suitable multivariate Gaussians provides a simple method to compute inverse matrices, which is particularly effective for large sparse systems [22]. It was used in [19] to evaluate the average number of sites visited by an unbiased random walker.

For $z < 1$ the matrix $\hat{R}(z)$ is positive definite, and so is its inverse. One can therefore evaluate elements of $\hat{R}(z)$ as averages

$$\hat{R}_{ij}(z) = \langle x_i x_j \rangle \quad (22)$$

over the multivariate Gaussian

$$\begin{aligned} P(\mathbf{x}) &= \frac{1}{Z} \exp \left[-\frac{1}{2} \mathbf{x}^T \hat{R}^{-1}(z) \mathbf{x} \right] \\ &= \frac{1}{Z} \exp \left[-\frac{1}{2} \sum_{ij \in \mathcal{V}} x_i \left(\delta_{ij} - z c_{ij} \sqrt{\frac{s(k_i)s(k_j)}{\Gamma_i \Gamma_j}} \right) x_j \right]. \end{aligned} \quad (23)$$

To proceed it is advantageous [24] to rescale variables $x_i/\sqrt{\Gamma_i} \rightarrow x_i$. Keeping the same symbols for the rescaled variables, we have

$$\begin{aligned} P(\mathbf{x}) &= \frac{1}{Z} \exp \left[-\frac{1}{2} \sum_{ij \in \mathcal{V}} x_i \left(\Gamma_i \delta_{ij} - z c_{ij} \sqrt{s(k_i)s(k_j)} \right) x_j \right] \\ &= \frac{1}{Z} \exp \left[-\frac{1}{2} \sum_{ij \in \mathcal{V}} c_{ij} \left(\frac{1}{2} (x_i^2 s(k_j) + x_j^2 s(k_i)) - z \sqrt{s(k_i)s(k_j)} x_i x_j \right) \right] \end{aligned} \quad (24)$$

for their joint distribution, where we have inserted the definition of the Γ_i in the second line and used the symmetry of the c_{ij} to express the resulting distribution in terms of an exponential of a manifestly symmetric quadratic form.

To evaluate the diagonal elements \hat{R}_{jj} in equation (15), one only needs single-site variances $P(\mathbf{x})$, for which only single-site marginals of $P(\mathbf{x})$ are needed. Following standard reasoning [19, 22, 24] one finds these as

$$P_i(x_i) = \int \left[\prod_{j \in \mathcal{V} \setminus i} dx_j \right] P(\mathbf{x}),$$

giving

$$P_i(x_i) \propto \int \left[\prod_{j \in \partial i} dx_j \right] \exp \left[- \sum_{j \in \partial i} \left(\frac{1}{2} (x_i^2 s(k_j) + x_j^2 s(k_i)) - z \sqrt{s(k_i) s(k_j)} x_i x_j \right) \right] P^{(i)}(\mathbf{x}_{\partial i}). \quad (25)$$

Here $P^{(i)}(\mathbf{x}_{\partial i})$ is the joint cavity marginal of the x_j on sites j which are neighbours of node i on the graph with node i missing. On a locally tree-like graph one has $P^{(i)}(\mathbf{x}_{\partial i}) \simeq \prod_{j \in \partial i} (P_j^{(i)}(x_j))$, where ∂i denotes the set of neighbours of i and the $P_j^{(i)}(x_j)$ are the single-site cavity marginals of x_j . Hence the integrals in equation (25) factor, and

$$P_i(x_i) \propto \prod_{j \in \partial i} \int dx_j \exp \left[- \frac{1}{2} (x_i^2 s(k_j) + x_j^2 s(k_i)) + z \sqrt{s(k_i) s(k_j)} x_i x_j \right] P_j^{(i)}(x_j). \quad (26)$$

Following the same line of reasoning for single-site cavity marginals $P_j^{(i)}(x_j)$, we have

$$P_j^{(i)}(x_j) \propto \prod_{\ell \in \partial j \setminus i} \int dx_\ell \exp \left[- \frac{1}{2} (x_j^2 s(k_\ell) + x_\ell^2 s(k_j)) + z \sqrt{s(k_j) s(k_\ell)} x_j x_\ell \right] P_\ell^{(j)}(x_\ell). \quad (27)$$

The system (27) of equations is self-consistently solved by Gaussians of the form

$$P_j^{(i)}(x_j) = \sqrt{\frac{\omega_j^{(i)}}{2\pi}} \exp \left[- \frac{1}{2} \omega_j^{(i)} x_j^2 \right], \quad (28)$$

with $\omega_j^{(i)} > 0$, entailing that the inverse cavity variances need (in the limit $z \rightarrow 1$) to satisfy the self-consistency equations

$$\omega_j^{(i)} = \sum_{\ell \in \partial_j \setminus i} \left[s(k_\ell) - \frac{s(k_j) s(k_\ell)}{\omega_\ell^{(j)} + s(k_j)} \right]. \quad (29)$$

With the $P_j^{(i)}(x_j)$ Gaussian, the single site marginals $P_i(x_i)$ are also Gaussian. Denoting inverse single-site variances by ω_i , we obtain these in terms of inverse cavity variances as

$$\omega_i = \sum_{j \in \partial i} \left[s(k_j) - \frac{s(k_i) s(k_j)}{\omega_j^{(i)} + s(k_i)} \right]. \quad (30)$$

Equations (29) and (30) generalise those obtained in [19] to cover general degree-biased random walk models.

Once equation (29) are solved for a given single instance of a graph, the inverse single-site marginals can be computed. When evaluating the search efficiency B according to equation (18) we need to recall that the ω_j are inverse single site variances of *rescaled* variables $x_j / \sqrt{\Gamma_j}$. We therefore have $\hat{R}_{jj} = \Gamma_j / \omega_j$, hence

$$B = \frac{1}{Y} \sum_{j \in \mathcal{V}} s(k_j) \omega_j \xi_j. \quad (31)$$

We will see in section 4 that, for sufficiently large systems, results obtained using the present approach agree very well with those of simulations. Before turning to results, however, we will first elaborate the theory for the limit of infinitely large systems.

3.2. Thermodynamic limit

In the thermodynamic limit, equation (29) can be interpreted as stochastic recursion relations for inverse variances of single-site cavity marginals. In what follows we will use these equations to obtain a system of self-consistency equations for the *distributions* of the inverse cavity variances for ensembles of random graphs in the configuration model class. It turns out that degree dependent families of such distributions are needed due to the node degrees appearing in equation (29). The resulting self-consistency equations for these distributions can be solved by a stochastic population dynamics algorithm [25]. The solution then determines the degree dependent distributions of inverse single site marginals needed to evaluate search efficiencies in the thermodynamic limit according to equation (31).

However, it turns out that the results of this approach cannot be directly compared to single large instance calculations or to simulations, which are usually performed by restricting attention to the (single) giant component of a graph, whereas standard random graph ensembles typically describe systems which—apart from the giant component—also contain finite clusters. In section 3.2.3 below we will therefore introduce the necessary modifications which will allow one to compute search efficiencies of random walkers restricted to the giant component of random graph ensembles.

3.2.1. Distributions of inverse cavity variances. The self-consistency equation (29) for the inverse cavity variances imply that $\omega_j^{(i)}$ depends on the node degree k_j of node j . Analogous degree dependences must therefore be expected for the $\omega_\ell^{(j)}$ appearing on the rhs of equation (29). In the thermodynamic limit, we therefore need to self-consistently determine an *entire family* $\{\tilde{\pi}_k(\tilde{\omega})\}_{k \geq 1}$ of *degree-dependent distributions* of inverse cavity variances $\tilde{\omega}$.

To evaluate these distributions, consider nodes of degree $k_j = k$ adjacent to some cavity site i . The probability $\tilde{\pi}_k(\tilde{\omega}) d\tilde{\omega}$ that $\omega_j^{(i)} \in (\tilde{\omega}, \tilde{\omega} + d\tilde{\omega}]$ for some pair (i, j) of adjacent vertices with $k_j = k$ is obtained by summing over the probabilities of realisations of the rhs of (29) for all j with $k_j = k$, which give a value in that range. Recall that $\omega_j^{(i)}$ has contributions from all vertices adjacent to j , except i . Thus, considering nodes of degree k adjacent to a cavity, we shall denote the collection of degrees of the $k - 1$ vertices (not including the cavity) connected to such a node by $\{k_\nu \geq 1\}_{k-1} = \{k_\nu \geq 1; \nu \in \{1, \dots, k-1\}\}$. In a configuration model, the probability of encountering such a collection of degrees of $k - 1$ neighbouring vertices is $\prod_{\nu=1}^{k-1} \frac{k_\nu}{c} p_{k_\nu}$, in which p_k denotes the probability of having a vertex of degree k in the graph, so that $\frac{k}{c} p_k$ is the probability that a randomly chosen neighbour of a node has degree k , with $c = \langle k \rangle$ denoting the mean degree. Introducing the shorthand notation $d\tilde{\pi}_{k_\nu}(\tilde{\omega}_\nu) = \tilde{\pi}_{k_\nu}(\tilde{\omega}_\nu) d\tilde{\omega}_\nu$, and defining

$$\Omega_{k-1}(\{\tilde{\omega}_\nu, k_\nu\} | k) = \sum_{\nu=1}^{k-1} \left(s(k_\nu) - \frac{s(k)s(k_\nu)}{\tilde{\omega}_\nu + s(k)} \right), \quad (32)$$

as a shorthand for a random realization of the rhs of (29) for $k_j = k$ we then have

$$\tilde{\pi}_k(\tilde{\omega}) = \sum_{\{k_\nu \geq 1\}_{k-1}} \left[\prod_{\nu=1}^{k-1} \frac{k_\nu}{c} p_{k_\nu} \right] \int \left[\prod_{\nu=1}^{k-1} d\tilde{\pi}_{k_\nu}(\tilde{\omega}_\nu) \right] \delta[\tilde{\omega} - \Omega_{k-1}(\{\tilde{\omega}_\nu, k_\nu\} | k)]. \quad (33)$$

In equation (33) $\delta[\dots]$ denotes the Dirac δ -distribution.

In a similar vein, the degree dependent distributions $\pi_k(\omega)$ of inverse single-site variances of the rescaled Gaussian variables x_j are obtained from equation (30) as

$$\pi_k(\omega) = \sum_{\{k_\nu \geq 1\}_k} \left[\prod_{\nu=1}^k \frac{k_\nu}{c} p_{k_\nu} \right] \int \left[\prod_{\nu=1}^k d\tilde{\pi}_{k_\nu}(\tilde{\omega}_\nu) \right] \delta[\omega - \Omega_k(\{\tilde{\omega}_\nu, k_\nu\}|k)]. \quad (34)$$

They can be evaluated once the solutions $\{\tilde{\pi}_k(\tilde{\omega})\}$ of equation (33) have been found. These distributions can be used to compute the search efficiency B from equation (31).

3.2.2. Search efficiencies. To evaluate search efficiencies, we rewrite equation (31) as

$$B = \frac{1}{Y/N} \left[\frac{1}{N} \sum_{j \in \mathcal{V}} s(k_j) \omega_j \xi_j \right], \quad (35)$$

thereby highlighting the fact that it is a ratio of two terms, which—in the thermodynamic limit $N \rightarrow \infty$ —can both be evaluated by appeal to the law of large numbers. Recalling from equation (14) that the normalisation constant Y appearing in the Perron–Frobenius eigenvector \mathbf{v}_1 of \hat{R} gives

$$\frac{Y}{N} = \frac{1}{N} \sum_{j \in \mathcal{V}} s(k_j) \Gamma_j,$$

we find that this results in

$$B = \frac{1}{\mathcal{N}} \sum_k p_k \left[s(k) \mathbb{E}[\omega|k] \mathbb{E}[\xi|k] \right] \quad (36)$$

as the limiting expression for the search efficiency. Here

$$\begin{aligned} \mathbb{E}[\omega|k] &= \int d\pi_k(\omega) \omega \\ &= \sum_{\{k_\nu \geq 1\}_k} \left[\prod_{\nu=1}^k \frac{k_\nu}{c} p_{k_\nu} \right] \int \left[\prod_{\nu=1}^k d\tilde{\pi}_{k_\nu}(\tilde{\omega}_\nu) \right] \Omega_k(\{\tilde{\omega}_\nu, k_\nu\}|k) \end{aligned} \quad (37)$$

by equation (34), and we have

$$\mathbb{E}[\xi|k] = \rho_h \frac{h(k)}{\langle h \rangle} \quad (38)$$

from equation (20), while

$$\mathcal{N} = \lim_{N \rightarrow \infty} \frac{Y}{N} = c \left[\sum_k \frac{k}{c} p_k s(k) \right]^2 \quad (39)$$

is the limiting value of the normalisation factor Y/N . We refer to appendix C for its evaluation.

The search efficiency B clearly has a natural decomposition in terms of contributions of vertices of different degree. It can be written as

$$B = \sum_{k \geq 1} p_k B_k, \quad (40)$$

where the k -dependent components B_k are given by

$$B_k = \frac{1}{\mathcal{N}} s(k) \mathbb{E}[\omega|k] \mathbb{E}[\xi|k]. \quad (41)$$

They denote the fraction of sites of degree k on which items are found per unit time in the course of an n -step degree-biased random walk.

The reasoning in the present section does not properly take into account the fact that simulations or single-instance cavity analyses are typically performed on graphs which consist of a *single component*, given that a random walker can only explore the graph component on which (s)he starts the search in the first place. If that component is one of the finite clusters, then only that finite cluster can be explored in the search so that the number of items found in a random walk will be finite, hence the efficiency of the search as defined by the number of items found per unit time in an n -step walk will tend to zero in the large- n limit. In the following section we will discuss the modifications of the theory necessary to take into account the fact that only random searches on the giant component of a random graph will give a non-zero contribution to the search efficiency B .

3.2.3. Isolating giant component contributions. As we have just indicated, any node belonging to one of the finite clusters would give a zero contribution to the search efficiency B in the thermodynamic limit, and only the nodes in the giant component are going to contribute to the result. It is therefore important to differentiate between the two and to be able to restrict results obtained for the search efficiency in the thermodynamic limit to contributions coming only from the giant component of the system.

In order to do this, we can follow [20], and supplement the recursions equation (29) for the inverse cavity variances and expression equation (30) for inverse single-site variances by analogous equations describing whether or not a site adjacent to a cavity belongs to the giant component of the system, and similarly whether a randomly selected site does or does not belong to the giant component.

This is achieved by introducing indicator variables n_i for each node i which take the value 1, if node i belongs to the giant component of a graph and 0, if it does not. In a similar vein, indicator variables $n_j^{(i)}$ are introduced to express whether a node j adjacent to a cavity site i does or does not belong to the giant component. For these we have

$$n_i = 1 - \prod_{j \in \partial i} (1 - n_j^{(i)}) \quad (42)$$

$$n_j^{(i)} = 1 - \prod_{\ell \in \partial j \setminus i} (1 - n_\ell^{(j)}) . \quad (43)$$

The first of these equations states that node i belongs to the giant component of the graph if at least one of its neighbours is connected to the giant-component through a path not involving i , whereas the second equation expresses the same fact for a site adjacent to the node i on the cavity graph from which node i is removed.

In the thermodynamic limit equation (43) can once more be thought of as stochastic recursions for random cavity indicator variables. For a node j of degree $k_j = k$ adjacent to a cavity node i we now seek to determine the joint probability $\tilde{\pi}_k(\tilde{\omega}, \tilde{n}) d\tilde{\omega}$ that the inverse cavity variance $\omega_j^{(i)}$ falls into the infinitesimal interval $(\tilde{\omega}, \tilde{\omega} + d\tilde{\omega}]$ and that the cavity indicator variable $n_j^{(i)}$ takes the value $n_j^{(i)} = \tilde{n} \in \{0, 1\}$. As for equation (33), this joint probability is obtained by summing over the probabilities of all realisations of the rhs of equations (29) and (43) which give a value of the inverse cavity variance in that prescribed range and a value \tilde{n} for the cavity indicator variable. This gives

$$\begin{aligned} \tilde{\pi}_k(\tilde{\omega}, \tilde{n}) &= \sum_{\{k_\nu \geq 1, \tilde{n}_\nu\}_{k-1}} \left[\prod_{\nu=1}^{k-1} \frac{k_\nu}{c} p_{k_\nu} \right] \int \left[\prod_{\nu=1}^{k-1} d\tilde{\pi}_{k_\nu}(\tilde{\omega}_\nu, \tilde{n}_\nu) \right] \delta[\tilde{\omega} - \Omega_{k-1}(\{\tilde{\omega}_\nu, k_\nu\}|k)] \\ &\times \delta_{\tilde{n}, 1 - \prod_{\nu=1}^{k-1} (1 - \tilde{n}_\nu)}. \end{aligned} \quad (44)$$

In a similar vein we obtain the joint distribution $\pi_k(\omega, n)$ for the inverse single-site variances ω_i and the single-site indicator variables n_i from the solution of equation (44) as

$$\begin{aligned} \pi_k(\omega, n) &= \sum_{\{k_\nu \geq 1, \tilde{n}_\nu\}_k} \left[\prod_{\nu=1}^k \frac{k_\nu}{c} p_{k_\nu} \right] \int \prod_{\nu=1}^k d\tilde{\pi}_{k_\nu}(\tilde{\omega}_\nu, \tilde{n}_\nu) \delta[\omega - \Omega_k(\{\tilde{\omega}_\nu, k_\nu\}|k)] \\ &\times \delta_{n, 1 - \prod_{\nu=1}^k (1 - \tilde{n}_\nu)}. \end{aligned} \quad (45)$$

The search efficiency B evaluated on the giant cluster can be written as

$$B = \frac{1}{Y_g/N_g} \left[\frac{1}{N_g} \sum_{j \in \mathcal{V}_g} s(k_j) \omega_j \xi_j \right], \quad (46)$$

where \mathcal{V}_g is the set of nodes in the giant cluster, N_g is the number of nodes in the giant cluster and $Y_g = \sum_{j \in \mathcal{V}_g} s(k_j) \Gamma_j$. In the thermodynamic limit, both the numerator and the denominator in this expression are once more evaluated by appeal to the law of large numbers. We will use a recent result of Tishby *et al* [26] about degree distributions conditioned on the giant component of random graphs to evaluate the denominator and use the $\pi_k(\omega, n)$ in (45) to compute conditional expectations of inverse single-site variances ω conditioned on degree *and* on nodes belonging to the giant cluster

$$\begin{aligned} \mathbb{E}[\omega|k, n=1] &= \int d\pi_k(\omega|1) \omega \\ &= \frac{1}{\rho} \sum_{\{k_\nu \geq 1, \tilde{n}_\nu\}_k} \left[\prod_{\nu=1}^k \frac{k_\nu}{c} p_{k_\nu} \right] \int \left[\prod_{\nu=1}^k d\tilde{\pi}_{k_\nu}(\tilde{\omega}_\nu, \tilde{n}_\nu) \right] \Omega_k(\{\tilde{\omega}_\nu, k_\nu\}|k) \\ &\times \delta_{1, 1 - \prod_{\nu=1}^k (1 - \tilde{n}_\nu)}, \end{aligned} \quad (47)$$

in which ρ is the probability of a randomly chosen vertex to belong to the giant cluster. This gives

$$B = \frac{1}{\mathcal{N}_g} \sum_{k \geq 1} p(k|1) \left[s(k) \mathbb{E}[\omega|k, n=1] \mathbb{E}(\xi|k) \right], \quad (48)$$

with $p(k|1)$ denoting the degree distribution *conditioned* on the giant cluster [26], and

$$\mathcal{N}_g = \frac{c}{\rho} \sum_{k, k'} \frac{k}{c} p_k \frac{k'}{c} p_{k'} s(k) s(k') [1 - (1 - \tilde{\rho})^{k'+k-2}] \quad (49)$$

giving the limiting value of Y_g/N_g ; its evaluation, following [26], is left to appendix C. In equation (49), $\tilde{\rho}$ denotes the probability of a random link pointing to nodes on the giant cluster. These quantities can be easily evaluated using standard generating function techniques.

As before, the expression for the search efficiency restricted to the giant component has a natural decomposition into contributions of vertices of different degrees, that is,

$$B = \sum_{k \geq 1} p(k|1) B_k . \quad (50)$$

Here, the degree distribution appearing in equation (40) is replaced by the degree distribution conditioned on the giant component, and we have

$$B_k = \frac{1}{\mathcal{N}_g} s(k) \mathbb{E}[\omega|k, n=1] \mathbb{E}[\xi|k] . \quad (51)$$

The self-consistency equation (44) for the $\tilde{\pi}_k(\tilde{\omega}, \tilde{n})$ which are needed to evaluate search efficiencies in the thermodynamic limit are very efficiently solved using a stochastic population dynamics algorithm. The new aspect in the present problem is that several such populations are needed to represent the $\tilde{\pi}_k(\tilde{\omega}, \tilde{n})$ for the different degrees k in the system.

3.3. Analytic results for random regular graphs

On random regular graphs, we have $p_k = \delta_{k,c}$. Hence there cannot be a non-trivial degree biased strategy, as the normalised matrix of transition probabilities is independent of the choice of $s(k) = s(c) = s$, and we are therefore looking at an unbiased random walk as the search strategy, and random hiding as the hiding strategy.

Given that all nodes (and all links) are equivalent in the thermodynamic limit, the solution of equations (32) and (33) is $\tilde{\pi}_c(\tilde{\omega}) = \delta(\tilde{\omega} - \bar{\omega})$, with $\bar{\omega}$ satisfying

$$\bar{\omega} = (c-1) \left[s - \frac{s^2}{\bar{\omega} + s} \right] . \quad (52)$$

The only non-zero solution to (52) is $\bar{\omega} = s(c-2)$, where $s = s(c)$. Using $\mathcal{N} = cs^2$ from equation (39), and inserting $\tilde{\pi}_c(\tilde{\omega}) = \delta(\omega - \bar{\omega})$ into equation (45) we have

$$\mathbb{E}[\omega|k] = \mathbb{E}[\omega|c] = cs \frac{c-2}{c-1} , \quad (53)$$

and thus

$$B = \rho_h \frac{c-2}{c-1} . \quad (54)$$

This result is independent of s as it should. The result was obtained in [19] from a single-instance cavity analysis of the case $\rho_h = 1$, and is also consistent with earlier findings of [27] for random walks on Bethe lattices of connectivity c .

3.4. Approximations

In what follows we will look at two approximate descriptions of the hide and seek problem.

The first is based on comparing the equilibrium distribution of the random walker executing the search with the distribution characterising the location of hidden items on the network. While this equilibrium type analysis does not actually provide us with an estimate of the search efficiency, it will allow us to find parameter settings for the strategy of the searcher which will optimise the search efficiency for a given hiding strategy.

The second approximation is based on a so-called non-backtracking assumption and it will actually produce approximate values for search efficiencies. For reasons to be described below, we expect these approximations to become quite accurate in the limit where most vertices of the system actually have large degrees.

3.4.1. An analysis using equilibrium distributions. The analysis in the present section is based on the observation that a random walker starting her walk on any randomly chosen site of a network will—after only a few steps of the walk—very quickly ‘forget’ about any specific properties of the starting vertex and start visiting different vertices of the system with probabilities given by the equilibrium probability of the random walker.

Let us denote by $q_s(k)$ the equilibrium probability of the random walker to visit a site of degree k , and by $q_h(k)$ the probability that a randomly selected site *with* an item hidden on it has degree k . Choosing parameters of the search strategy in such a way that q_s is as close as possible to q_h should then provide a good heuristic to optimise the efficiency of a search strategy.

From equation (20) we have

$$q_h(k) = p_k \frac{h(k)}{\langle h \rangle}, \tag{55}$$

with $\langle h \rangle = \sum_k p_k h(k)$ for the conditional probability that a site has degree k given that an item is hidden on it. In a similar fashion we have

$$q_s(k) = \sum_{i \in \mathcal{V}} \pi_i \delta_{k_i, k} = p_k \frac{ks(k)}{\langle ks \rangle} \tag{56}$$

for the probability that a random walker in equilibrium finds herself on a site of degree k .

A measure of the similarity of q_s and q_h is given by the Kullback–Leibler (KL) divergence between them, which is given by

$$\text{KL}(q_s || q_h) = \sum_{k=1}^{\infty} q_s(k) \log \left[\frac{q_s(k)}{q_h(k)} \right] = \sum_{k=1}^{\infty} \frac{ks(k)}{\langle ks \rangle} p_k \log \left[\frac{ks(k)}{\langle ks \rangle} \frac{\langle h \rangle}{h(k)} \right]. \tag{57}$$

See e.g. [28], chapter 2. Minimising the KL divergence over any parameters characterising the search strategy is then expected to provide a good indication of the parameter setting for the most efficient search strategy within the parameterised family of strategies under consideration.

For power-law search $s(k) = k^\alpha$ pitted against the power-law hiding strategy $h(k) = k^\beta$, the minimisation of the KL divergence can be done analytically, and it leads to

$$\alpha = \beta - 1 \tag{58}$$

for the exponent of the most efficient search strategy. We will see in the results section that the result is far from exact. The main reason is, of course, that the number of marked sites visited at least once is what matters for the search efficiency whereas the total frequency of visits (including repeated visits) to sites is the quantity determining the equilibrium distribution.

3.4.2. A non-backtracking approximation. On a network in which degrees are typically large, the probability of an unbiased random walker to return to the site from which she transitioned to the site she is currently on becomes *small*. This is because the probability of choosing any particular neighbour as the target of the next step, and thereby the probability of retracing the last step is inversely proportional to the degree of the site the random walker currently finds herself on, which is therefore small for a site with a large number of neighbours.

For a degree-biased random walker, this effect will persist unless the degree-bias in the transition probabilities is extremely strong (e.g. such that the walker almost always goes to the neighbouring site with the largest degree: one can easily convince oneself that in such a

situation there can be configurations of neighbouring sites at which the random walker could be trapped once it hits such a set of sites.).

Thus, assuming that back-tracking events are rare, a non-backtracking approximation can be formulated as follows. Denote by $S(\xi, n)$ the average number of items found in an n -step walk. We suppress the index of site from which the walker started out its search, as we have learned above that for large n the search efficiency will be independent of the starting site. Assume that the random walker at the n th step of her walk visits the site j , coming from a site i , which is adjacent to j . If site j is visited for the first time, the only chance to find additional items in the next step is *not to backtrack* on the previous step, but to visit sites $\ell \in \partial j \setminus i$. Taking averages over the sites, using the equilibrium distribution π_i to give the probability that the walker found herself on site i in step $n - 1$, and assuming that j and its neighbours (apart from i) are being visited for the first time, we obtain

$$S(\xi, n + 1) = S(\xi, n) + \sum_{i \in \mathcal{V}} \pi_i \sum_{j \in \partial i} W_{ij} \sum_{\ell \in \partial j \setminus i} W_{j\ell} \xi_\ell. \quad (59)$$

This recursion is easily solved. Taking $S(\xi, 0) = 0$ as the initial condition, we get

$$S(\xi, n) = Bn, \quad (60)$$

with

$$B = \sum_{i \in \mathcal{V}} \pi_i \sum_{j \in \partial i} W_{ij} \sum_{\ell \in \partial j \setminus i} W_{j\ell} \xi_\ell. \quad (61)$$

Using the transition probabilities (7) and the resulting expression (10) for the equilibrium distribution we obtain

$$B = \frac{1}{Y} \sum_{i \in \mathcal{V}} s(k_i) \Gamma_i \sum_{j \in \partial i} \frac{s(k_j)}{\Gamma_i} \sum_{\ell \in \partial j \setminus i} \frac{s(k_\ell)}{\Gamma_j} \xi_\ell. \quad (62)$$

Repeating the line of reasoning that lead to the expression (36) for search efficiencies in the thermodynamic limit, we can evaluate equation (62) in this limit. The resulting expression is

$$B = \frac{c}{\mathcal{N}} \sum_k \frac{k}{c} p_k s(k) \sum_{k'} \frac{k'}{c} p_{k'} s(k') \sum_{\{k_\nu\}_{\nu=1}^{k'-1}} \left[\prod_{\nu=1}^{k'-1} \frac{k_\nu}{c} p_{k_\nu} s(k_\nu) \right] \sum_{\nu=1}^{k'-1} \frac{s(k_\nu) \mathbb{E}[\xi | k_\nu]}{s(k) + \sum_{\nu=1}^{k'-1} s(k_\nu)}. \quad (63)$$

We shall see in section 4.2 below that this approximation is remarkably efficient even for systems with moderate values of their mean degree.

4. Results

We now turn to results. We will evaluate search efficiencies for large finite systems using (i) the single-instance cavity approach described in section 3.1 and (ii) numerical simulations. Search efficiencies in the thermodynamic limit will be analysed using (iii) the methods described in section 3.2. We will use these three approaches to explore how various search strategies fare against a range of hiding strategies. We shall find that there is an excellent agreement between results obtained using simulations and those obtained using either the single-instance cavity method or the method designed for the large system limit, provided the projection to the giant component described in section 3.2.3 is used in the thermodynamic limit, and finite single instances are sufficiently large.

As mentioned in section 2.2, we cover several functional forms for the degree bias of both hiding and searching strategies, namely, power-law, exponential, and logarithmic strategies. We evaluate search efficiencies across the spectrum of functional forms used to describe hiding and searching strategies and—for given functional forms—across parameter ranges characterising them.

Finally we investigate search efficiencies for various graph types, including Erdős–Rényi graphs and scale-free graphs, and we assess the quality of our approximate approaches by comparing them with exact results.

4.1. Validating the theory

We begin by validating the theoretical approaches described in section 3, by comparing their results with those of stochastic simulations. We do this initially for complete occupancy $\xi_i \equiv 1$, where the number of different sites visited by a walker is a measure of network exploration efficiency (rather than search efficiency).

Random networks of a sufficiently large size are generated, and n -step degree biased random walks starting from a randomly chosen vertex on the giant cluster are simulated. The number of different sites visited is recorded. As noted in section 3 that number is for sufficiently large n expected to be independent of the starting vertex and $S(n) \sim Bn$ for $1 \ll n \ll N$. We determine the exploration efficiency B by averaging over many realisations of the random walk and over many realisations of random graphs in the given ensemble. Alternatively, we compute the exploration efficiency B directly using the cavity method, averaging the results over the same set of graphs. The cavity method requires to solve equation (29) for single large instances. We found that for all cases considered in the present paper, this can very effectively be done by simple forward iteration.

For finite single instances we find that graph sizes $N = 6000$ were sufficient to compare simulation results with those obtained via the cavity method on the one hand side, and with the thermodynamic limit results on the other hand side. All finite single instance results shown below will therefore have been obtained for systems of this size. The optimal n range for which the behaviour of $S(n)$ can be fitted by a linear law has been determined by minimising χ^2 in linear regression. Figure 1 shows the results of simulations and confirms the linear behaviour of $S(n)$ for intermediate n . We found that $40 \lesssim n \lesssim 230$ was an optimal range for the linear fit for this system, but observed that slightly narrower fitting ranges were required for other degree biases. From the simulations we determine $B = 0.716\,727 \pm 0.000\,203$. This compares well with the analysis of B evaluated directly via the cavity approach, which gives $B = 0.716\,789 \pm 0.000\,210$.

In figure 2, we compare the results of simulations with those obtained from the cavity analysis for degree-biased random walkers with power-law degree bias $s(k) = k^\alpha$ for a range of α values between $\alpha = -5$ and $\alpha = +5$, and we observe very good agreement between the two. The cavity method can therefore be safely used as a substitute of random walk simulations for computing exploration and search efficiencies. In figure 2 and below the symbols show the measured B values, while the connecting lines are guides to the eye. Errors of both simulation and cavity results are estimated to be $\mathcal{O}(10^{-4})$ for the exploration efficiencies presented in figure 2, and $\mathcal{O}(10^{-5})$ for the search efficiencies, so error bars are mostly significantly smaller than the symbols indicating B values. The same is true for results presented in the remainder of this paper.

The α dependence of B can be understood by noting that very negative α will force the walker to spend most of her time at low degree sites, which are themselves surrounded by low

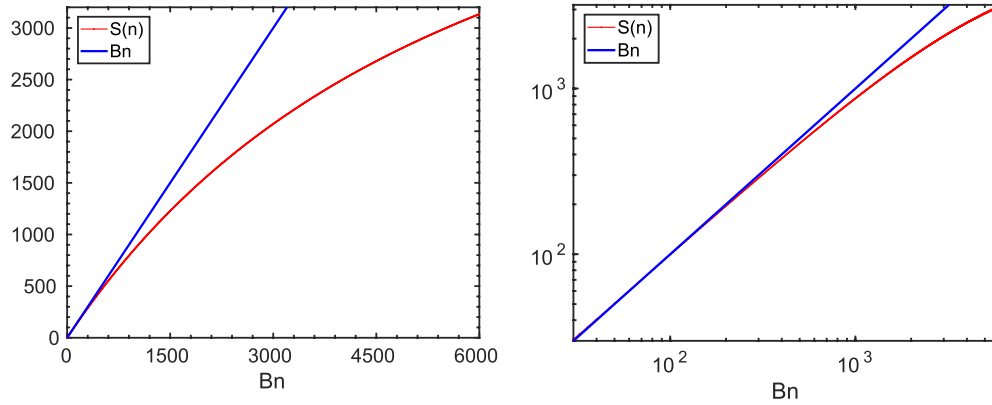


Figure 1. Behaviour of $S(n)$ for a degree-biased random walker with degree bias following a power-law $s(k) = k^\alpha$ with $\alpha = 1$. The left panel displays both $S(n)$ (lower curve) and Bn (upper curve), with $B \simeq 0.716872$ determined from simulations, as functions of Bn . The right panel shows the same results on a double-logarithmic plot. The behaviour of $S(n)$ is well described by the linear law for not too large n . For larger n there is a clear crossover to sub-linear behaviour due to finite size effects. Results were obtained for Erdős–Rényi graphs of mean degree $c = 4$. Simulations were performed on the giant component of graphs whose original size was $N = 6000$. For $c = 4$ the giant component occupies a fraction $\rho \simeq 0.98$ of the entire system. Results of simulation runs are averaged over $N_s = 2000$ random graph realisations.

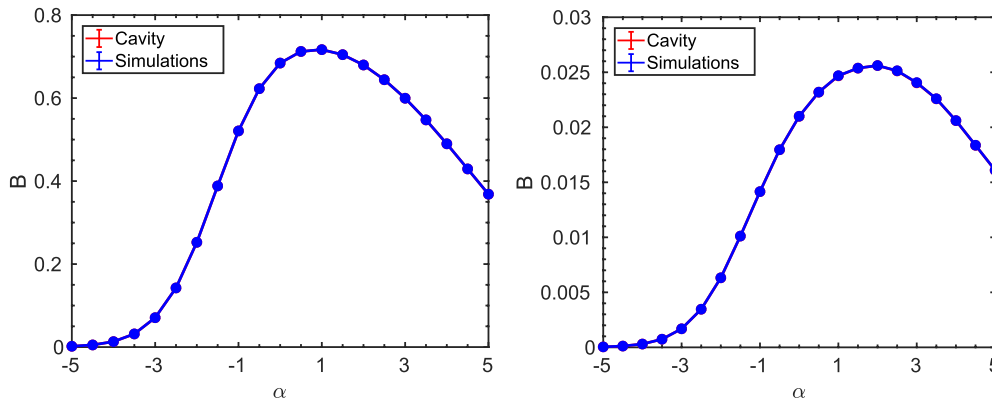


Figure 2. Left panel: network exploration efficiency B of a power-law degree biased random walk $s(k) = k^\alpha$ on Erdős–Rényi graphs of mean degree $c = 4$. Right panel: search efficiency of a power-law degree biased random walk computed for power-law degree biased hiding with $h(k) = k$ for the case where a fraction $\rho_h = 0.025$ of sites have an item hidden on them. The connecting line is a guide to the eye. On the scale of the figure, results obtained from the cavity method are indistinguishable from those obtained from random walk simulations. The cavity results were obtained for giant components of systems of size $N = 6000$, averaged over $N_s = 2000$ random graphs.

degree sites, i.e. at the end of dangling chains in the graph, whereas very large positive α will entail that the walker is very likely to be found on sites with very high degrees that are themselves surrounded by high-degree nodes. Both extremes would prevent efficient exploration of the network, so large values of B are expected at intermediate α .

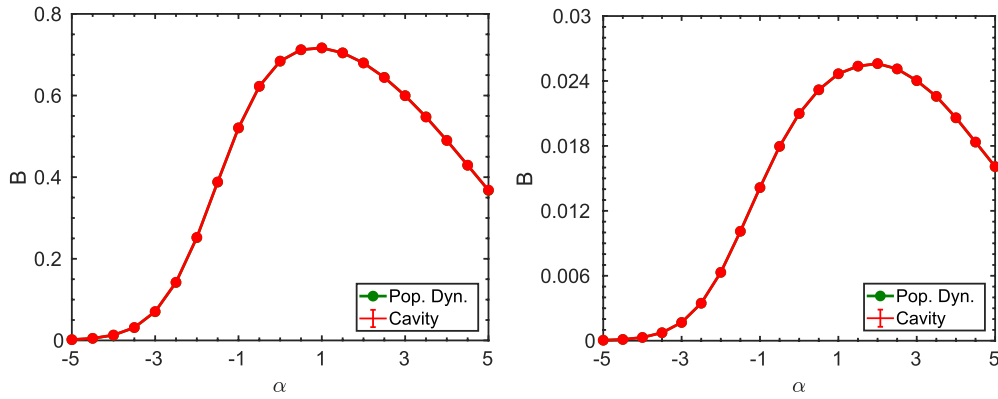


Figure 3. Comparison of cavity and thermodynamic limit results for power-law biased random walk $s(k) = k^\alpha$ on Erdős-Rényi networks of mean degree $c = 4$. Left panel: network exploration efficiency computed for $\xi_i \equiv 1$. Right panel: search efficiency computed for degree biased power-law hiding with $h(k) = k$ for the case where a fraction $\rho_h = 0.025$ of sites has an item hidden on it. The cavity results were obtained for giant components of systems of size $N = 6000$, averaged over $N_s = 2000$ random graphs.

In figure 3, finally we compare results of the single-instance cavity approach performed on the giant component of random graphs with those obtained using the theory for the thermodynamic limit. We see in figure 3 that there is an excellent agreement between results obtained via averaging cavity results over single large problem instances and results obtained in the thermodynamic limit, provided the projection onto the giant component described in section 3.2.3 is performed.

If one were to perform simulations by randomly selecting a starting vertex from the *entire system*, the starting vertex would belong to the giant component with probability ρ , whereas with probability $1 - \rho$ it would belong to one of the finite clusters of the system. The contribution of the latter to search and exploration efficiencies is zero, so one would expect average efficiencies for the entire system to obey

$$B = \rho B_g + (1 - \rho) B_f = \rho B_g, \quad (64)$$

with B_g and B_f denoting search and exploration efficiencies corresponding to the giant component and the finite clusters of the system, respectively.

Naïvely applying the thermodynamic limit theory of sections 3.2.1 and 3.2.2 does *not* produce this result (nor even $B \propto B_g$ with a proportionality constant that is independent of the search-strategy). The reason for this is that one of the key assumptions underlying the evaluation of search and exploration efficiencies along the lines described in section 3, viz. the fact that the Perron-Frobenius eigenvalue of the transition matrix is unique, ceases to be valid when the system contains several clusters and the random walk transition matrix is thus decomposable.

4.2. Hide and seek

We now look at pitting different hiding and searching strategies against each other. The main questions to be answered are concerned with identifying best search strategies (within a given family), when pitted against hiding strategies (again within a given family). Conversely, one

might wish to identify the most efficient hiding strategy, when pitted against given search strategies.

Before presenting those results, let us point out though that the probability of hiding items in any of the degree biased hiding strategies is according to equation (20) proportional to the overall fraction ρ_h of sites with an item hidden on them. It is therefore expected, and explicitly borne out by equations (36) and (48) that search efficiencies in the large system limit will be proportional to ρ_h . We verify this explicitly in figure 4 by displaying the ratio $B(\rho_h)/\rho_h$ as a function of search parameter α for power-law search pitted against a degree biased hiding strategy of the form $h(k) = k$. Note that $B(\rho_h)/\rho_h > 1$ for the optimal α value, implying that the searcher is able to exploit the degree bias of the hider to locate hidden items more effectively than expected by the fraction of sites with items hidden on them. Unless stated otherwise we have in what follows chosen $\rho_h = 0.025$ for the density of hidden items.

In figure 5 we investigate the search efficiency of power-law search (left panel) and of exponential search (right panel), when pitted against power-law hiding. We observe that there are optimal values of parameters of the search strategy which depend on the exponent characterising the power-law hiding strategy. Optimal search efficiencies are comparable in both cases, though matched functional forms for the degree bias of hiding and searching generally perform slightly better than unmatched forms. The range of reasonably effective search parameters is narrower for the exponential family. This is easily understood as, for a given value of the bias parameter, exponential bias is generally more efficient in creating heterogeneity of transition rates than power-law bias.

Figure 6 displays the efficiencies of power-law search (left panel) and exponential search (right panel), when set against logarithmic hiding of the form $h(k) = \log(1 + \beta k^{\gamma_h})$ with $\gamma_h = 1$. In this figure we use the convention that $\beta = 0$ is meant to refer to unbiased hiding. Note that in both cases the searcher's efficiency is always larger for degree-biased logarithmic hiding than for the unbiased hiding strategy with $\beta = 0$.

Figure 7 illustrates the decomposition of exploration efficiencies according to equation (51) into contributions B_k of vertices of different degree k encountered in a degree biased walk with power-law degree bias of the form $s(k) = k^\alpha$. Peak positions indicating the degrees of sites which give the largest contributions to network exploration efficiencies are increasing functions of the bias parameter α of the random walker. Peak heights vary with α , with the largest peak height corresponding to the optimal exploration bias $\alpha \simeq 1$ as observed in figure 2.

Finally we address the question of the influence of the graph type on search or exploration efficiencies. In figure 8 we look at exploration efficiencies of a degree biased random walker on Erdős–Rényi, scale-free, and random regular graphs, as well as on a real-world network—a symmetrized version of the Gnutella peer-to-peer file sharing network [29]. For the random regular graph, any form of degree bias is clearly ineffective and the exploration efficiency must obviously be independent of the value of a formal bias parameter, as indeed confirmed by the results. Results also confirm the analytic prediction $B = 2/3$ obtained in equation (54) for the $c = 4$ system. The Gnutella network has an average degree of $c = 4.819$, slightly higher than the mean degree of the synthetic networks, yet close enough to make for a meaningful comparison. For this real-world network, we have performed additional tests to verify that results of the cavity analysis agree with those of numerical simulations. We found the agreement to be better than fractions of a per-cent. Results indicate that the degree bias is more effective in enhancing the exploration efficiency in the scale-free graph and the Gnutella network than in the Erdős–Rényi graph, which is presumably due to the greater heterogeneity of the degree distributions in the scale-free system and the Gnutella network. For the latter though, the slightly higher mean connectivity may have further contributed to improved exploration efficiency at most α values.

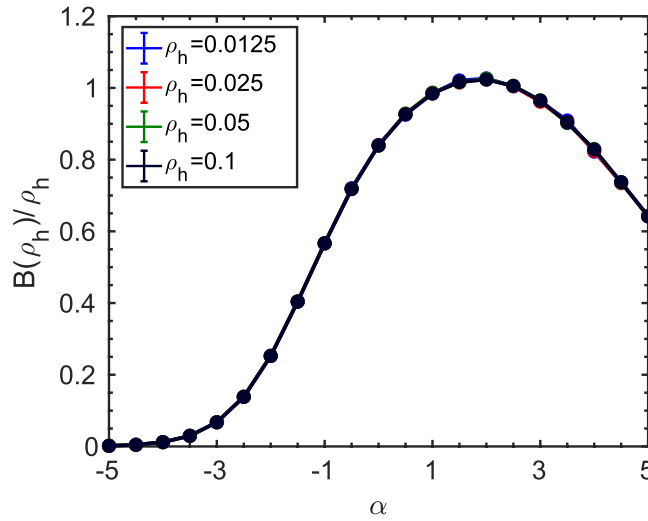


Figure 4. Efficiency of power-law search with $s(k) = k^\alpha$ when set against power-law hiding of the form $h(k) = k^\beta$ with $\beta = 1$. Shown are the ratios $B(\rho_h)/\rho_h$ for various values of ρ_h in the allowed range defined by equation (21), obtained by the single instance cavity method for the giant component of Erdős–Rényi graphs with $c = 4$ and $N = 6000$, averaged over $N_s = 2000$ instances. Curves lie on top of each other, verifying the expected proportionality.

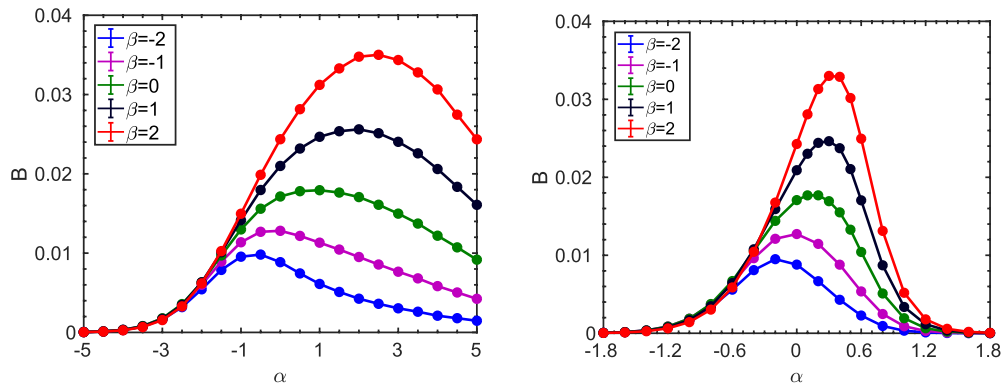


Figure 5. Efficiency of power-law search with $s(k) = k^\alpha$ (left panel) and of exponential search with $s(k) = e^{\alpha k}$ (right panel) as functions of α , when set against power-law hiding of the form $h(k) = k^\beta$ for various β , and $\rho_h = 0.025$. In both panels, curves from bottom to top correspond to increasing values of the bias parameter β of the hiding strategy. Shown are single instance cavity results for the giant component of Erdős–Rényi graphs with $c = 4$ and $N = 6000$, averaged over $N_s = 2000$ instances.

We now turn to approximations. In figure 9, values of the KL divergence (57) as a function of the degree bias of the searcher are displayed together with the search efficiencies for the examples of power-law search set against random and power-law hiding. While low values of the KL divergence are a reasonable qualitative predictor for high search efficiencies, the relation is not quantitative. In fact the minimum of the KL divergence occurs at a value of α which is approximately $\Delta\alpha \simeq 2$ below the value for which the search efficiency is maximised.

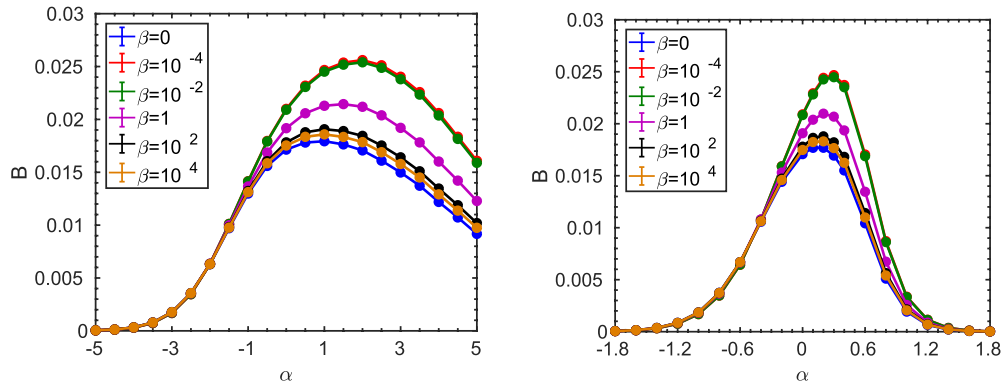


Figure 6. Efficiency of power-law search with $s(k) = k^\alpha$ (left panel) and exponential search with $s(k) = e^{\alpha k}$ (right panel) set against logarithmic hiding of the form $h(k) = \log(1 + \beta k)$ for various β , and $\rho_h = 0.025$, with $\beta = 0$ meant to refer to unbiased random hiding. In both panels, curves from bottom to top correspond to increasing values of the bias parameter β of the hiding strategy. Shown are single instance cavity results for the giant component of Erdős–Rényi graphs with $c = 4$ and $N = 6000$, averaged over $N_s = 2000$ instances.

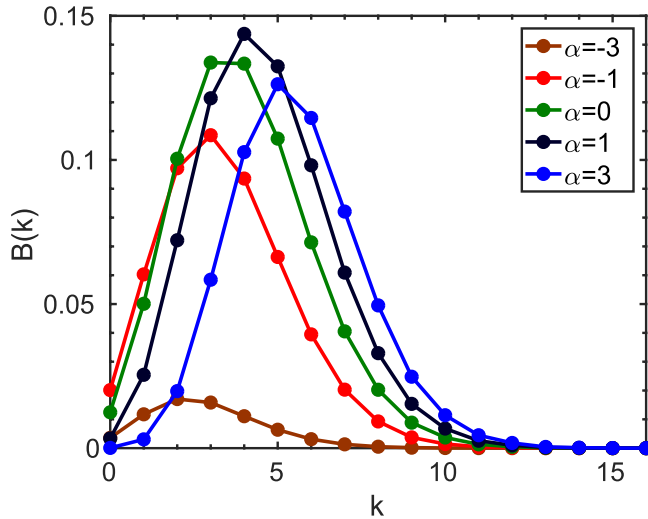


Figure 7. Decomposition of the network exploration efficiency B into k -dependent contributions for a range of bias parameters of the random walker with power-law bias of the form $s(k) = k^\alpha$. Shown are result obtained using population dynamics for the giant component of Erdős–Rényi graphs with mean degree $c = 4$.

A discrepancy between the bias values which minimise the KL divergence and which maximise the values of the search efficiency is of course not unexpected, as the KL divergence is based on equilibrium considerations whereas the search efficiency is a manifestly non-equilibrium measure, as it does not account for (the frequency of) multiple visits of any given site, which is a characteristic equilibrium property.

In figure 10 we investigate the power of the non-backtracking approximation for network exploration and search efficiencies. We expect this approximation to be efficient in networks

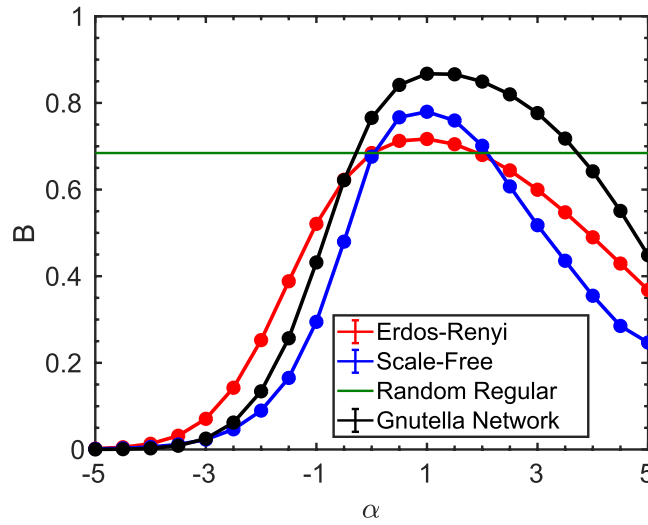


Figure 8. Comparison of network exploration efficiencies for four different graph types using the cavity method. Parameters are $N = 6000$, and $c = 4$ for Erdős–Rényi and regular random graphs; for the scale-free graph we chose $\gamma = 2.65$, with $k_{\min} = 2$, $k_{\max} = 400$ giving a mean connectivity $c = 3.905$. Finally the Gnutella network is a peer-to-peer file sharing network [29], consisting of $N = 36\,682$ nodes, from which we have created an undirected version by symmetrising the links. Its average degree is $c = 4.819$. The degree distribution of the Gnutella network exhibits two regimes with distinct power law behaviours, viz. $1 \leq k \leq 9$ where $p_k \propto k^{-1.74}$, and $11 \leq k \leq 40$ where $p_k \propto k^{-4.91}$. Curves with peak heights from bottom to top correspond to the Erdős–Rényi, the scale-free and the Gnutella network, respectively. As expected there is no effect of degree bias on the exploration efficiency for the random regular graph.

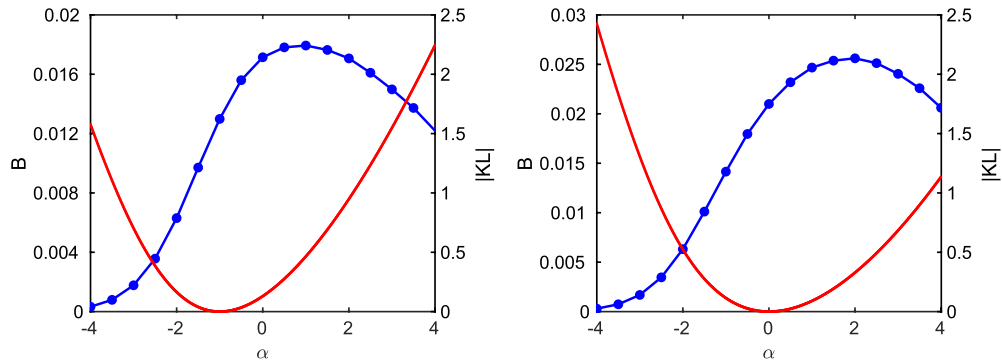


Figure 9. Search efficiency and KL divergence displayed as functions of the bias parameter α for power-law search $s(k) = k^\alpha$, set against random hiding (left panel) and power-law hiding $h(k) = k^\beta$, with $\beta = 1$ (right panel). In both panels, values of search efficiencies are displayed on the left axis, and those for KL divergences on the right axis. Search efficiencies were obtained using cavity for Erdős–Rényi graphs of size $N = 6000$, with $c = 4$ and $\rho_h = 0.025$, averaged over $N_s = 2000$ samples.

in which there are very few sites with low degrees, such as large mean degree Erdős–Rényi graphs. Our results show that the non-backtracking approximation is highly efficient qualitatively in that it predicts optimal search and exploration parameters very accurately already for a $c = 4$ Erdős–Rényi graph. While the actually predicted search and exploration efficiencies

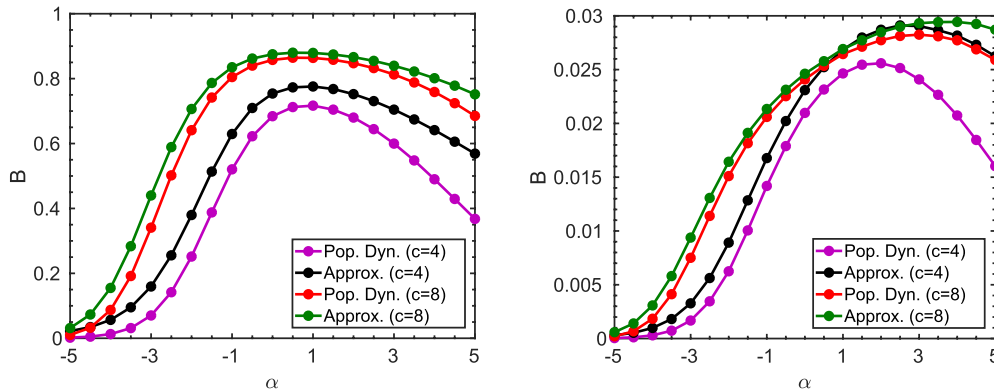


Figure 10. Network exploration efficiency of a degree-biased random walker with degree bias following a power-law $s(k) = k^\alpha$ as a function of the bias parameter α (left panel). Efficiency of power-law search with $s(k) = k^\alpha$, set against power-law hiding $h(k) = k^\beta$, with $\beta = 1$ as a function of the bias parameter α (right panel). In both panels we compare results obtained via population dynamics for the thermodynamic limit with those of the non-backtracking approximation described in section 3.4. The upper pair of curves in the left panel was computed for Erdős–Rényi graphs of mean degree $c = 8$, whereas the lower pair of curves shows results for Erdős–Rényi graphs of mean degree $c = 4$. The same trend is observed in the right panel, except for a small range of positive α values where the non-backtracking approximation predicts larger values of exploration efficiencies for the $c = 4$ system than for the $c = 8$ system.

are for such a low mean degree system still $\mathcal{O}(25\%)$ off the mark, the approximation improves markedly across the entire α range studied in the system with a (still moderate) mean connectivity $c = 8$. The approximation is therefore remarkably powerful, given that it is fairly straightforward, and in fact conceptually and technically *much* simpler than the full solution.

5. Summary and discussion

We have studied the efficiency of random search strategies to locate items hidden on a subset of vertices of complex networks, using a random walk framework. We assumed that items are hidden according to stochastic, degree biased strategies. In order to evaluate search efficiencies we adapt a result of De Bacco *et al* [19] in which the average number of different vertices of a complex network visited by random walker performing an unbiased n -step random walk is computed, generalising their work by considering more general degree biased transition probabilities.

We use the cavity method to compute diagonal elements of resolvents needed for the evaluation of network exploration and search efficiencies for large single problem instances. We also derive results for search efficiencies valid in the thermodynamic limit $N \rightarrow \infty$ of infinite system size. This requires the solution of a degree dependent family of non-linear integral equations for inverse cavity variances. Their solution is obtained using a suitably adapted version of the population dynamics algorithm of Mézard and Parisi [25].

It turns out that the naïve derivation, based on simply re-interpreting finite-instance self-consistency equations for inverse cavity variances as stochastic recursions in the thermodynamic limit does *not* accurately capture results valid for the giant component in the thermodynamic limit. The theory needs to be supplemented by degrees of freedom capturing

whether sites do or do not belong to the giant cluster, as proposed in [20] in the context of sparse random matrix spectra.

With this amendment, we find that results obtained using the cavity method for the giant components of large single instances of size $N = 6000$ are already in excellent agreement with those obtained for the giant cluster in the thermodynamic limit, and both are in turn in excellent agreement with those of stochastic random walk simulations.

We found that search and network exploration efficiencies have a natural decomposition in terms of degrees of sites contributing to the overall result, and we provided such a deconvolution of the network exploration efficiency by degree for the example of a random walker with a power-law degree bias. It is fairly easy to see that this type of deconvolution could be carried out beyond degree, thereby identifying local environments (such as degree and the collection of degrees in the first coordination shell) that are most conducive to network exploration or search.

We have looked at various parameterised families of degree biased search algorithms and degree biased hiding strategies, namely power-law, exponential, and logarithmic families, as described in table 1. Whatever the hiding strategy, we find—for each family of search strategies—a unique intermediate value of the parameter characterising the search strategy that can be considered as optimal in the sense that it maximises the search efficiency. An analogous statement can be made for the network exploration efficiency. Qualitatively this could be understood by recalling that extreme values of the search parameter tended to imply that a degree biased random walker would spend most of her time at very low or very high degree sites, with both extremes not being conducive to efficient search or exploration.

We verified that search efficiencies were proportional to the density of items hidden in the network, and we observed that normalised search efficiencies $B(\rho_h)/\rho_h$ could be larger than 1 if there was a sufficiently strong degree bias in the hiding strategy which could be exploited to boost search.

Pitting matched and unmatched functional forms of hiding and search strategies against each other, we always observed that the optimal search strategy in matched families slightly outperformed the most efficient search strategy when the functional forms of hiding and search strategies were unmatched.

We also used equilibrium dynamics consideration to locate efficient values for search parameters by looking at the Kullback–Leibler distance between the distribution of degrees the random walker visits in equilibrium and the distribution of degrees with items hidden on them. The other approximation we looked into is a so-called non-backtracking approximation. It is based on the intuition that random walks on networks experience an effective drift away from the starting vertex, which becomes very effective if vertices typically exhibit large degrees. In such a situation one can evaluate search efficiencies assuming that—locally—every non-backtracking step explores unseen parts of the network. We expect this approximation to be efficient in situations where networks have few low degree sites, and we demonstrated that it was surprisingly effective on Erdős–Rényi graphs even with modest mean degrees.

When comparing network exploration efficiencies for different graph types, we observed that, in an intermediate α range of power-law bias of the random walker, exploration on scale-free graphs is more efficient than on Erdős–Rényi graphs, as a degree biased walker can make more efficient use of the large degree of heterogeneity of vertex degrees in scale-free graphs than of the rather small degree of heterogeneity present on Erdős–Rényi networks. Large positive and large negative biases tend to localise the walker (near a set of nodes or sets of nodes) with either high or low degree vertices, leading to a reduction in exploration of search efficiencies. For the random regular graph, any degree bias is ineffective hence the search efficiency is found to be independent of the degree bias α as expected.

The hide and seek scenarios so far considered either did or did not have an item hidden on a vertex. It is easy to see that efficiencies can be computed with graded values attached to the items hidden on each site.

Our analysis has been restricted to computing *average* search and exploration efficiencies. Clearly, from a security point of view discussed in the introduction it would be interesting to compute distributions of search efficiencies, in order to assess, for instance, the likelihood of conducting unusually efficient, or unusually in-efficient searches. This would be particularly relevant if one were to insist on locating all items hidden in the net and would thus have determine the cover-time [15], though not for the entire network but for a specified subset of vertices. Such questions are for the time being outside the reach of our methods. They could, of course, always be addressed using simulations.

We have so far not dealt with proper game theoretical questions such as with the existence and characterisation of Nash equilibria in the present problem, or with the possibility of agents *learning* efficient search strategies, either on the fly or in repeated instances of the game. Analogous problems can be posed for the hider, who could update their hiding strategy in repeated instances of the game, by observing the efficiency of any strategy used by the seeker.

We intend to address some of these questions in future publications.

Acknowledgments

The authors acknowledge funding by the UK Engineering and Physical Sciences Research Council (EPSRC) through the Centre for Doctoral Training ‘Cross Disciplinary Approaches to Non-Equilibrium Systems’ (CANES, Grant Nr. EP/L015854/1). Insightful discussions with Peter McBurney are also gratefully acknowledged.

Appendix A. Random walk analysis

In what follows we briefly summarise the key elements of the derivation of equation (15) which forms the basis of the evaluation of search efficiencies, closely following [19].

The average number $S_i(\xi, n)$ of marked sites visited in an n -step random walk, given by (5), is expressed in terms of the probabilities $H_{ij}(n)$ of visiting node j at least once in the first n time steps when starting at node i . The $H_{ij}(n)$ can in turn be decomposed according to the time m of the last visit to j as

$$H_{ij}(n) = \sum_{m=0}^n G_{ij}(m) q_{jj}(n-m), \quad (\text{A.1})$$

in which $q_{jj}(n-m)$ denotes the probability for a walker who started at node j not to return to node j in $n-m$ steps, and $G_{ij}(m) = (W^m)_{ij}$ is the m -step transition probability from i to j . The convolution structure of the above expression entails

$$\hat{H}_{ij}(z) = \hat{G}_{ij}(z) \hat{q}_{jj}(z) \quad (\text{A.2})$$

for its z -transform. The $q_{jj}(n)$ are in turn related with first passage probabilities $F_{jj}(n)$ via

$$q_{jj}(n-1) - q_{jj}(n) = F_{jj}(n),$$

from which, with $q_{jj}(0) = 1$ and $F_{jj}(0) = 0$, one obtains

$$\hat{q}_{jj}(z) = \frac{1 - \hat{F}_{jj}(z)}{1 - z}. \quad (\text{A.3})$$

From

$$G_{ij}(n) = \delta_{ij}\delta_{n0} + \sum_{m=0}^n F_{ij}(m)G_{jj}(n-m) \quad (\text{A.4})$$

finally one gets

$$\hat{G}_{ij}(z) = \frac{1}{1 - \hat{F}_{jj}(z)}, \quad (\text{A.5})$$

and thus

$$\hat{H}_{ij}(z) = \frac{1}{1-z} \frac{\hat{G}_{ij}(z)}{\hat{G}_{jj}(z)}, \quad (\text{A.6})$$

from which the z -transform of the number of items found in an n -step walk is obtained as

$$\hat{S}_i(\boldsymbol{\xi}, z) = \frac{1}{1-z} \sum_{j \in \mathcal{V}} \frac{\hat{G}_{ij}(z)}{\hat{G}_{jj}(z)} \xi_j. \quad (\text{A.7})$$

This is equation (6) in section 2.

Appendix B. Spectral analysis

To evaluate $\hat{S}_i(\boldsymbol{\xi}, z)$ further one uses equation (11) and the spectral decomposition (13),

$$\hat{R}(z) = \frac{\mathbf{v}_1 \mathbf{v}_1^T}{1-z} + \sum_{\nu=2}^N \frac{\mathbf{v}_\nu \mathbf{v}_\nu^T}{1-z\lambda_\nu} \equiv \frac{\mathbf{v}_1 \mathbf{v}_1^T}{1-z} + \hat{C}(z), \quad (\text{B.1})$$

of $\hat{R}(z)$. Using equation (14) for the components of the Perron Frobenius eigenvector and $\hat{G}_{ij}(z) = \hat{R}_{ij}(z)$ one has

$$\hat{S}_i(\boldsymbol{\xi}, z) = \frac{1}{1-z} \sum_{j \in \mathcal{V}} \left[\frac{s(k_j)\Gamma_j}{\hat{R}_{jj}(z)Y(1-z)} + \frac{\sqrt{\frac{s(k_j)\Gamma_j}{s(k_i)\Gamma_i}} \hat{C}_{ij}(z)}{\frac{s(k_j)\Gamma_j}{Y(1-z)} + \hat{C}_{jj}(z)} \right] \xi_j, \quad (\text{B.2})$$

where, following [19], we have used the spectral decomposition (B.1) of $\hat{R}_{jj}(z)$ in the denominator of the second contribution within the square brackets in (B.2). Noting that $Y \propto N$, the second contribution can be argued to be negligible in the limit of large system size $N \rightarrow \infty$ and $z \rightarrow 1$ (in this order; see [19], whereas the first contribution gives

$$\hat{S}_i(\boldsymbol{\xi}, z) \sim \frac{1}{(1-z)^2 Y} \sum_{j \in \mathcal{V}} \frac{s(k_j)\Gamma_j}{\hat{R}_{jj}} \xi_j, \quad z \rightarrow 1, \quad (\text{B.3})$$

i.e. equation (15), where it is assumed that

$$\hat{R}_{jj} = \lim_{z \rightarrow 1} \lim_{N \rightarrow \infty} \hat{R}_{jj}(z) = \lim_{z \rightarrow 1} \hat{C}_{jj}(z) \quad (\text{B.4})$$

exists.

Appendix C. Normalization factors

Here we describe the evaluation of the normalisation constants which appear in the expressions for the search and exploration efficiencies B in equations (35) and (46).

Looking at the normalisation constant Y/N in equation (35),

$$\frac{Y}{N} = \frac{1}{N} \sum_{i \in \mathcal{V}} s(k_i) \Gamma_i, \quad (\text{C.1})$$

we evaluated it in the thermodynamic limit $N \rightarrow \infty$ as a sum of averages by appeal to the law of large numbers (LLN). This gives $Y/N \rightarrow \mathcal{N}$ as $N \rightarrow \infty$, with

$$\mathcal{N} = \sum_k p_k s(k) \mathbb{E} \left[\Gamma_i \mid k_i = k \right], \quad (\text{C.2})$$

which is further evaluated as

$$\begin{aligned} \mathcal{N} &= \sum_k p_k s(k) \mathbb{E} \left[\sum_{j \in \partial_i} \mathbb{E} \left[s(k_j) \mid k_i = k \right] \right] \\ &= \sum_k p_k s(k) \left[k \sum_{k'} p(k'|k) s(k') \right] = \sum_k p_k s(k) \left[k \sum_{k'} \frac{k'}{c} p'_k s(k') \right], \end{aligned} \quad (\text{C.3})$$

where we have used the fact that, for configuration model networks, the probability $p(k'|k)$ that a site with degree k' is adjacent to a site with degree k does not depend on k , and is given by $p(k'|k) = \frac{k'}{c} p'_k$ in the last step. This is equation (39).

The evaluation of Y_g/N_g in equation (46) follows the same pattern, except for two crucial modifications. First, the degree distribution p_k used above needs to be replaced by the degree distribution $p(k|n=1)$ conditioned on the giant cluster. Second, the giant cluster of a configuration model network is *not* a configuration model itself, so the conditional probability $p(k'|k, n=1)$ that a vertex of degree k' is adjacent to a degree k site on a giant cluster does depend on the degree k . We can use results of Tishby *et al* [26] who recently provided a comprehensive analysis of the micro-structure of the giant component of configuration model networks, including the two ingredients needed here.

We have

$$\frac{Y_g}{N_g} = \frac{1}{N_g} \sum_{i \in \mathcal{V}_g} s(k_i) \Gamma_i, \quad (\text{C.4})$$

where N_g is the size of the giant component. In the thermodynamic limit $N_g = \rho N \rightarrow \infty$ we have $Y_g/N_g \rightarrow \mathcal{N}_g$ by the LLN, where

$$\begin{aligned} \mathcal{N}_g &= \sum_k p(k|n=1) s(k) \mathbb{E} \left[\sum_{j \in \partial_i} \mathbb{E} \left[s(k_j) \mid k_i = k, n = 1 \right] \right] \\ &= \sum_k p(k|n=1) s(k) \left[k \sum_{k'} p(k'|k, n=1) s(k') \right]. \end{aligned} \quad (\text{C.5})$$

Using

$$p(k|n=1) = \frac{1}{\rho} \left[1 - (1 - \tilde{\rho})^k \right] p_k \quad (\text{C.6})$$

and

$$p(k'|k, n=1) = \left[\frac{1 - (1 - \tilde{\rho})^{k'-1} (1 - \tilde{\rho})^{k-1}}{1 - (1 - \tilde{\rho})^k} \right] \frac{k'}{c} p'_k \quad (\text{C.7})$$

from [26], in which ρ is the percolating fraction, and $\tilde{\rho}$ is the probability that a neighbour of a randomly selected vertex is on the giant component of the system, we obtain

$$\mathcal{N}_g = \frac{c}{\rho} \sum_{k,k'} \frac{k}{c} p_k \frac{k'}{c} p'_k s(k) s(k') \left[1 - (1 - \tilde{\rho})^{k'+k-2} \right]. \quad (\text{C.8})$$

This is equation (49).

ORCID iDs

Shubham Pandey  <https://orcid.org/0000-0003-1415-9728>

Reimer Kühn  <https://orcid.org/0000-0001-8030-1951>

References

- [1] Alpern S and Gal S 2006 *The Theory of Search Games and Rendezvous* vol 55 (Berlin: Springer)
- [2] Alpern S, Fokkink R, Gasieniec L, Lindelauf R and Subrahmanian V S 2013 *Search Theory* (Berlin: Springer)
- [3] Alpern S, Morton A and Papadaki K 2011 Patrolling Games *Oper. Res.* **59** 1246–57
- [4] Tsai J, Yin Z, Kwak J, Kempe D, Kiekintveld C and Tambe M 2010 Urban security: Game-theoretic resource allocation in networked domains *24th AAAI Conf. on Artificial Intelligence*
- [5] Tsai J, Nguyen T H and Tambe M 2012 Security games for controlling contagion *26th AAAI Conf. on Artificial Intelligence*
- [6] Vaněk O, Yin Z, Jain M, Bošanský B, Tambe M and Pěchouček M 2012 Game-theoretic resource allocation for malicious packet detection in computer networks *Proc. 11th Int. Conf. on Autonomous Agents and Multiagent Systems* vol 2 (Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems) pp 905–12
- [7] Chapman M, Tyson G, McBurney P, Luck M and Parsons S 2014 Playing hide-and-seek: an abstract game for cyber security *Proc. 1st Int. Workshop on Agents and Cyber Security* (ACM) p 3
- [8] Tong H, Faloutsos C and Pan J-Y 2006 Fast random walk with restart and its applications *Proc. 6th Int. Conf. on Data Mining* (Washington, DC: IEEE Computer Society) pp 613–22
- [9] Bray A J and Rodgers G J 1988 Diffusion in a sparsely connected space: a model for glassy relaxation *Phys. Rev. B* **38** 11461–70
- [10] Moreno Y, Pastor-Satorras R and Vespignani A 2002 Epidemic outbreaks in complex heterogeneous networks *Eur. Phys. J. B* **26** 521–9
- [11] Newman M E J 2002 Spread of epidemic disease on networks *Phys. Rev. E* **66** 016128
- [12] Grady L 2006 Random walks for image segmentation *IEEE Trans. Pattern Anal. Mach. Intell.* **28** 1768–83
- [13] Lovász L 1993 Random walks on graphs: a survey *Combinatorics, Paul Erdős is Eighty* vol 2 (Budapest: Bolyai Mathematical Society) pp 1–46
- [14] Noh J D and Rieger H 2004 Random walks on complex networks *Phys. Rev. Lett.* **92** 118701
- [15] Cooper C and Frieze A 2005 The cover time of random regular graphs *SIAM J. Discrete Math.* **18** 728–40
- [16] Avin C, Koucký M and Lotker Z 2008 How to explore a fast-changing world (cover time of a simple random walk on evolving graphs) *Int. Coll. on Automata, Languages, and Programming* (Berlin: Springer) pp 121–32
- [17] Metzler R, Redner S and Oshanin G 2014 *First-Passage Phenomena and their Applications* (Singapore: World Scientific)

- [18] Levin D A, Peres Y and Wilmer E L 2017 *Markov Chains and Mixing Times* vol 107 (Providence, RI: American Mathematical Society)
- [19] De Bacco C, Majumdar S N and Sollich P 2015 The average number of distinct sites visited by a random walker on random graphs *J. Phys. A: Math. Theor.* **48** 205004
- [20] Kühn R 2016 Disentangling giant component and finite cluster contributions in sparse random matrix spectra *Phys. Rev. E* **93** 042110
- [21] Watts D J and Strogatz S H 1998 Collective dynamics of ‘small-world’ networks *Nature* **393** 440
- [22] Rogers T, Castillo I, Kühn R and Takeda K 2008 Cavity approach to the spectral density of sparse symmetric random matrices *Phys. Rev. E* **78** 031116
- [23] Edwards S F and Jones R C 1976 The eigenvalue spectrum of a large symmetric random matrix *J. Phys. A: Math. Gen.* **9** 1595–603
- [24] Kühn R 2015 Spectra of random stochastic matrices and relaxation in complex systems *Eur. Phys. Lett.* **109** 60003
- [25] Mézard M and Parisi G 2001 The Bethe lattice spin glass revisited *Eur. Phys. J. B* **20** 217–33
- [26] Tishby I, Biham O, Katzav E and Kühn R 2018 Revealing the micro-structure of the giant cluster in random graph ensembles *Phys. Rev. E* **97** 042318
- [27] Hughes B D and Sahimi M 1982 Random walks on the Bethe lattice *J. Stat. Phys.* **29** 781–94
- [28] Cover T A and Thomas J A 2006 *Elements of Information Theory* 2nd edn (Hoboken, NJ: Wiley)
- [29] Leskovec J and Sosič R 2016 Snap: a general-purpose network analysis and graph-mining library *ACM Trans. Intell. Syst. Technol.* **8** 1