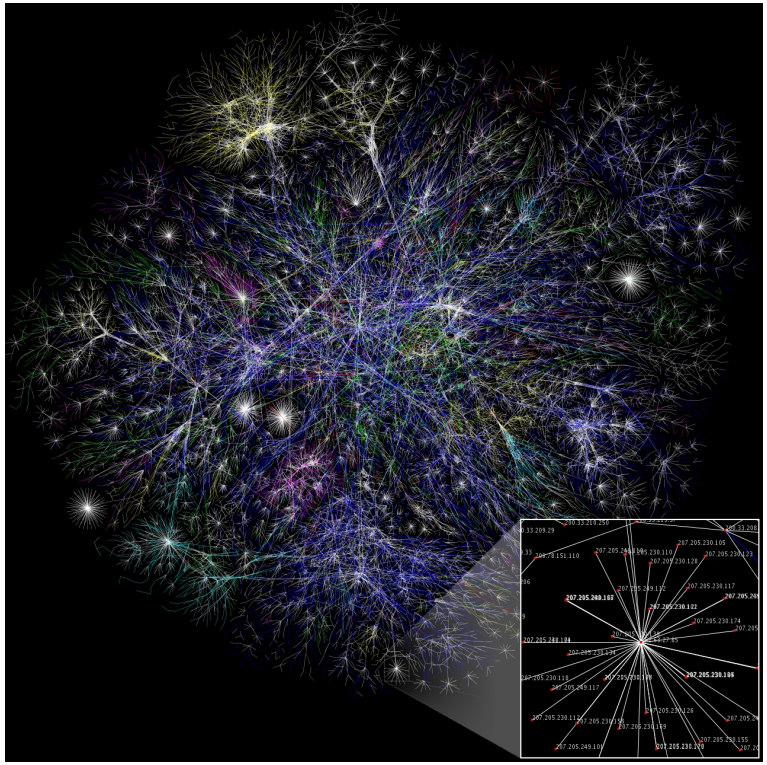
A Random Walk Perspective on Search and Network Exploration

Shubham Pandey and Reimer Kühn

Disordered Systems Group
Department of Mathematics, King's College London

Random Matrix Theory: Foundations and Applications, Krakow, 30 Apr-03 May, 2019



[www/opte.org:Internet 2007]

Outline

Outline

Exploration and Search

- Goal: to evaluate exploration and search efficiencies of random walkers.
- Search domains: complex networks
- Efficiency measured in terms of
  - average number $S_i(n)$ of different sites visited in an n step walk starting at vertex i (exploration).
  - average number $S_i(n|\xi)$ of different sites j with items hidden on them ($\xi_j = 1$) visited in an n step walk starting at vertex i (search).
- Applications
  - assess efficiency of diffusive spread of information in networks (exploration)
  - locate viruses hidden in computer networks (search)
  - assess efficiency of web-crawlers used to update search results for search engines (search)
- Note: search context has game-theoretic aspects (strategies of hider and seeker)

Outline

Number of Different Sites Visited by a Random Walker

- Known results for $n \gg 1$:
  - For d dimensional lattices and Bethe lattices: $S_i(n)$ independent of i.

$$S(n) \simeq \begin{cases} \sqrt{8n/\pi} & , \ d = 1 \\ \pi n / \ln n & , \ d = 2 \\ B(d) n & , \ d \geq 3 \\ \frac{c-2}{c-1} n & , \ \text{degree c BL} \end{cases}$$

[Dvoretzky and Erdős (1951), Vineyard (1963), Montroll and Weiss (1965), Hughes and Sahimi (1982)]

  - For random graphs in the configuration model class

$$S(n) \simeq Bn \ ,$$

where B depends on graph type, i.e. the degree distribution (p_k). Results generally not available in closed form. [De Bacco, Majumdar and Sollich (2015)]

Number of Different Sites Visited by a Random Walker

- Evaluation following [De Bacco, Majumdar Sollich (2015)]
  - Express $S_i(n)$ in terms of probabilities $H_{ij}(n)$ of visiting j at least once in n step walk starting in i,
$$S_i(n) = \sum_j H_{ij}(n)$$
  - Decompose $H_{ij}(n)$ according to time m of last visit to j
$$H_{ij}(n) = \sum_{m=0}^{n} G_{ij}(m) q_{jj}(n-m)$$
    with
    - $q_{jj}(n-m)$ denoting the probability for a walker starting at j at not to return to node j in $n-m$ steps,
    - $G_{ij}(m) = (W^m)_{ij}$ denoting the m-step transition probability from $i \to j$.
  - In terms of z-transforms, $\hat{f}(z) = \sum_{n=0}^{\infty} f(n) z^n$,
$$\hat{H}_{ij}(z) = \hat{G}_{ij}(z)\, \hat{q}_{jj}(z)$$

Number of Different Sites Visited by a Random Walker

- Evaluation (continued)
  - Exploit
    - relation between $q_{jj}(n)$ and first passage probabilities $F_{jj}(n)$

$$q_{jj}(n-1) - q_{jj}(n) = F_{jj}(n) \qquad \Leftrightarrow \qquad \hat{q}_{jj}(z) = \frac{1-\hat{F}_{jj}(z)}{1-z}$$

    - relation between $G_{ij}(n)$ and first passage probabilities $F_{ij}(n)$

$$G_{ij}(n) = \delta_{ij}\delta_{n0} + \sum_{m=0}^{n} F_{ij}(m) G_{jj}(n-m) \ ,$$

      or, in terms of z transforms,

$$\hat{G}_{jj}(z) = \frac{1}{1-\hat{F}_{jj}(z)} \ ,$$

    - entailing

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

  - Thus finally

$$\hat{S}_i(z) = \frac{1}{1-z} \sum_j \frac{\hat{G}_{ij}(z)}{\hat{G}_{jj}(z)}$$

Number of Different Sites Visited by a Random Walker

- Evaluate $S_i(n)$ for $n \gg 1$ from $z \to 1$-asymptotics of

$$\hat{S}_i(z) = \frac{1}{1-z} \sum_j \frac{\hat{G}_{ij}(z)}{\hat{G}_{jj}(z)}$$

- Requires knowledge of

$$\hat{G}(z) = \left[\mathbb{1} - zW\right]^{-1} .$$

- We will consider general degree-biased random walks

$$W_{ij} = W_{i \to j} = \frac{c_{ij} s(k_j)}{\Gamma_i} , \qquad \text{with} \qquad \Gamma_i = \sum_j c_{ij} s(k_j)$$

where $c_{ij} = c_{ji} = 1$, if i and j are connected, and $c_{ij} = 0$, if not.

Spectral Analysis

- To analyse $\hat{G}(z) = \left[\mathbb{I} - zW\right]^{-1}$, note that W satisfies a a detailed balance condition with the equilibrium distribution

$$\pi_i = \frac{1}{Y}\Gamma_i s(k_i)\ , \qquad \text{with} \qquad Y = \sum_j s(k_j)\Gamma_j\ .$$

- Use this to express $\hat{G}(z)$ in terms of a symmetric matrix

$$\hat{G}(z) = D^{-1/2}\hat{R}(z)D^{1/2}\ , \qquad \text{with} \qquad D = \mathrm{diag}(\Gamma_i s(k_i))\ ,$$

with

$$\hat{R}(z) = \left[\mathbb{I} - z\mathcal{W}\right]^{-1}\ ,$$

in which $\mathcal{W} = D^{1/2}WD^{-1/2}$ is symmetric.

Spectral Analysis

- Evaluation of $\hat{G}(z)$ via spectral decomposition of $\hat{R}(z)$

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

  where we have isolated the contribution of the Perron-Frobenius eigenvector $\mathbf{v}_1$ of $\mathcal{W}$ corresponding to $\lambda_1 = 1$ of W, with entries $v_{1,i} = \sqrt{\pi_i}$.

- For irreducible W, one has $|\lambda_\nu| < 1$ for $\nu \neq 1$, entaililing that the contribution of $\hat{C}(z)$ becomes negligible in the $z \to 1$ limit, giving

$$\hat{S}_i(z) \sim \frac{1}{(1-z)^2 Y} \sum_j \frac{s(k_j)\Gamma_j}{\hat{R}_{jj}} \ , \qquad z \to 1 \ ,$$

  with

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

Exploration and Search Efficiency

- The $1/(1-z)^2$ divergence of $\hat{S}_i(z)$, tranlates into

$$S_i(n) \sim B\,n\,, \qquad n \gg 1\,,$$

independently of i, with exploration efficiency

$$B = \frac{1}{Y}\sum_j \frac{s(k_j)\Gamma_j}{\hat{R}_{jj}}\,,$$

with the $N \to \infty$ limit assumed to be taken in this expression.

- This is trivially generalized to give a search efficiency

$$B(\xi) = \frac{1}{Y}\sum_j \frac{s(k_j)\Gamma_j}{\hat{R}_{jj}}\xi_j\,.$$

- Note: the $\hat{R}_{jj}$ still need to be evaluated! $\Rightarrow$ Cavity method.

Hide & Seek

- Investigate search efficiencies for various strategies of placing hidden items in the graph.
- Consider degree biased hiding strategies

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

in which ρ_h is the fraction of sites carrying a hidden item, h is a function of the degree, and $\langle h \rangle = \sum_k p_k h(k)$.

- Hiding and search strategies and their paremetrisations

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	$h(k) = \log(1 + \beta k^{\gamma_h})$	$s(k) = \log(1 + \alpha k^{\gamma_s})$

- Investigate matched and unmatched scenarios.

Outline

Cavity Method

- Non-trivial component in evaluation of search efficiencies

$$B(\xi) = \frac{1}{Y} \sum_j \frac{s(k_j)\Gamma_j}{\hat{R}_{jj}} \xi_j \ .$$

  is evaluation of the $\hat{R}_{jj}$

- Recall

$$\hat{R}_{jj}(z) = \left[\mathbb{I} - z\mathcal{W} \right]^{-1}$$

- Following Edwards and Jones (1976), can write

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

  where $\langle \ldots \rangle$ is an average over the multivariate Gaussian

$$P(\mathbf{x}) = \frac{1}{Z} \exp\left[-\frac{1}{2} \mathbf{x}^T \hat{R}^{-1}(z) \mathbf{x} \right]$$

Cavity Method

- Rescaling variables $x_i/\sqrt{\Gamma_i} \to x_i$, can write this as

$$P(\mathbf{x}) = \frac{1}{Z}\,\exp\left[-\frac{1}{2}\sum_{i,j} c_{ij}\left(\frac{1}{2}\left[x_i^2 s(k_j) + x_j^2 s(k_i)\right] - z\sqrt{s(k_i)s(k_j)}\,x_i x_j\right)\right]$$

- Only single site marginals are needed for the evaluation of $\hat{R}_{jj}(z)$
- Cavity approach [Rogers, Perez-Castillo, Takeda, RK (2008), De Bacco, Majumdar, Sollich (2015)]
- Need single site marginals for distribution of the form

$$P(\mathbf{x}) = \frac{1}{Z}\,\exp[-H(\mathbf{x})]\,, \qquad \text{with} \qquad H(\mathbf{x}) = \frac{1}{2}\sum_{i,j} V(x_i, x_j)$$

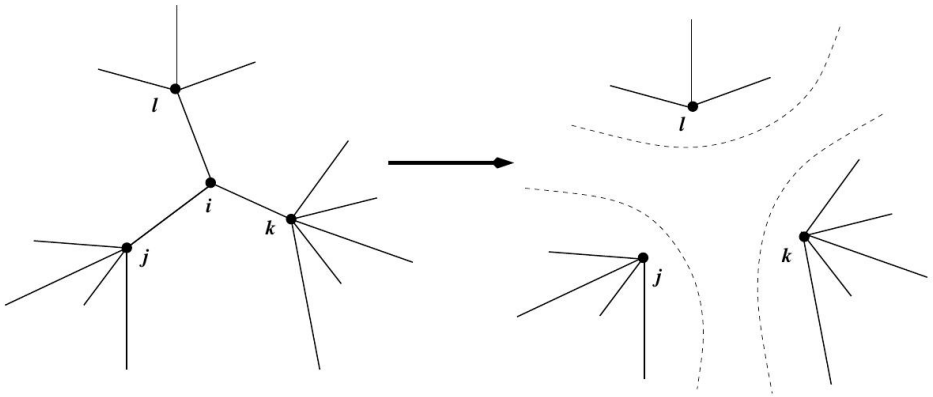
Cavity Method

- For all i, can write

$$H(\mathbf{x}) = \sum_{j \in \partial i} V(x_i, x_j) + H^{(i)}(\mathbf{x}_{\partial i}, \mathbf{x}_{\partial^2 i})$$

- Visualize decomposition

$H(\mathbf{x})$



$H^{(i)}(\mathbf{x}_{\partial i}, \mathbf{x}_{\partial^2 i})$

Left: Full system described by $H(\mathbf{x})$. Right: Cavity graph described by $H^{(i)}(\mathbf{x}_{\partial i}, \mathbf{x}_{\partial^2 i})$

Cavity Method

- Single site marginal thus

$$
\begin{aligned}
P_i(x_i) &= \frac{1}{Z}\int \mathrm{d}\mathbf{x}_{\partial i}\mathrm{d}\mathbf{x}_{\partial^2 i}\,\exp\Big[-\sum_{j\in\partial i} V(x_i,x_j) - H^{(i)}(\mathbf{x}_{\partial i},\mathbf{x}_{\partial^2 i})\Big] \\
&\propto \int \mathrm{d}\mathbf{x}_{\partial i}\,\exp\Big[-\sum_{j\in\partial i} V(x_i,x_j)\Big] P^{(i)}(\mathbf{x}_{\partial i})
\end{aligned}
$$

- On locally tree-like graph

$$
P_i(x_i) \propto \prod_{j\in\partial i}\int \mathrm{d}x_j\,\exp\Big[-V(x_i,x_j)\Big] P_j^{(i)}(x_j)
$$

- By same line of reasoning

$$
P_j^{(i)}(x_j) \propto \prod_{\ell\in\partial j\setminus i}\int \mathrm{d}x_\ell\,\exp\Big[-V(x_j,x_\ell)\Big] P_\ell^{(j)}(x_\ell) \qquad (*)
$$

Cavity Method

- In the present case

$$V(x_i, x_j) = c_{ij}\left(\frac{1}{2}\left[x_i^2 s(k_j) + x_j^2 s(k_i)\right] - z\sqrt{s(k_i)s(k_j)}\, x_i x_j\right)$$

  is harmonic.

- Entails that cavity-recursion $(*)$ is self-consistently solved by Gaussians

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

- Cavity-recursion $(*)$ gives (as $z \to 1$)

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

- Solve iteratively for given large single instances.

Cavity Method

- Single-site marginals $P_i(x_i)$, too, are Gaussian with inverse variances

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

- Search efficiencies in terms of these (need to undo $x_i/\sqrt{\Gamma_i} \to x_i$ transformation).

$$B(\xi) = \frac{1}{Y} \sum_i s(k_i)\omega_i \xi_i .$$

Thermodynamic Limit

- In infinite system limit interpret cavity-recursion $(*)$ as stochastic recursion for inverse variances of single-site cavity marginals.
- Use $(*)$ to derive system of self-consistency equations for the *distributions* of the inverse cavity variances for ensembles of random graphs in the configuration model class.
- Note:
  - due to the structure of $(*)$ need degree dependent families of such distributions
  - for general random graph ensembles, need to formulate and solve these projecting to the giant component of these networks, (Perron-Frobenius eigenvector is otherwise non-unique
  - do this by combining $(*)$ with stochastic recursions for indicator variables of the percolation problem [RK, Phys. Rev. E (2016)]

$$
\begin{aligned}
n_i &= 1 - \prod_{j \in \partial i} \left(1 - n_j^{(i)}\right) \\
n_j^{(i)} &= 1 - \prod_{\ell \in \partial j \setminus i} \left(1 - n_\ell^{(j)}\right) .
\end{aligned}
$$

Thermodynamic Limit

- Recursion equations for joint distributions of inverse cavity variances and cavity indicator variables

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

with

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

- From solution (obtained using population dynamics) get joint distributions of inverse single site variances and GC indicator variables

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

Search Efficiencies in the Thermodynamic Limit

- Putting things together we get

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

- Here
  - $p(k|1)$ is the degree distribution conditioned on the giant cluster

$$p(k|1) = \frac{1}{\rho} \left[1 - (1-\tilde{\rho})^k \right] p_k$$

  with ρ denoting the fraction of sites on the giant cluster, and $\tilde{\rho}$ the probability that a random link connects to a site on the giant cluster.

  [I. Tishby, O. Biham, E. Katzav, and RK, Phys. Rev. E (2018)]

  - We also have

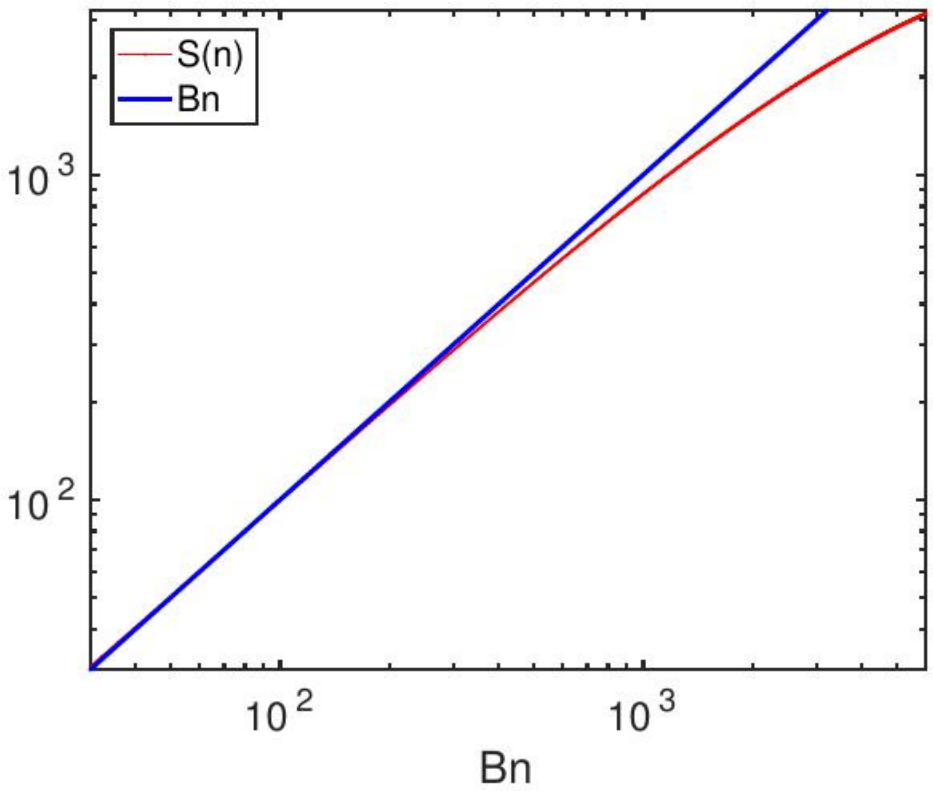
$$\mathbb{E}(\xi|k) = \rho_h \frac{h(k)}{\langle h \rangle} \qquad \mathbb{E}[\omega|k, n=1] = \frac{1}{\rho} \int \mathrm{d}\pi_k(\omega, 1)\, \omega$$

  - and

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

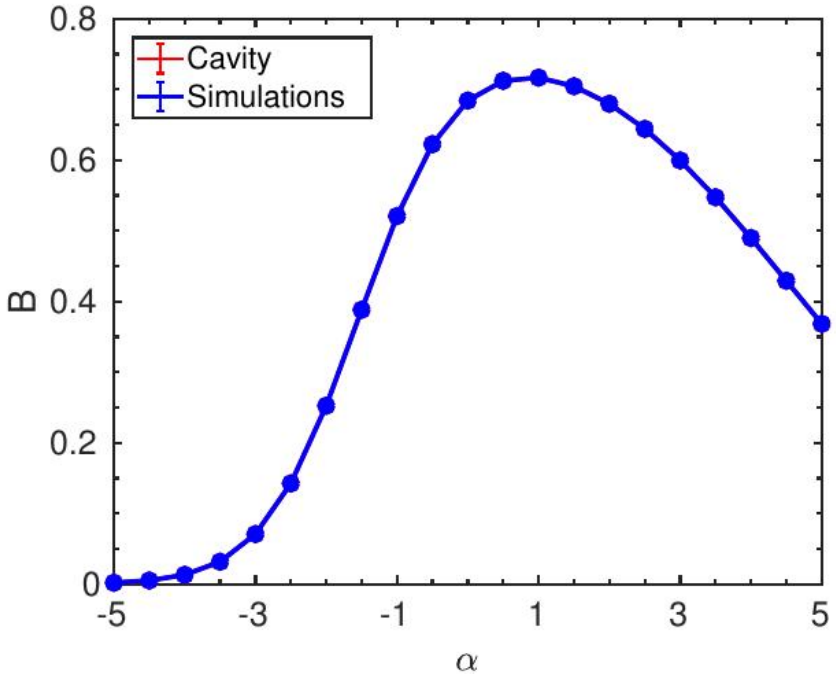
Outline

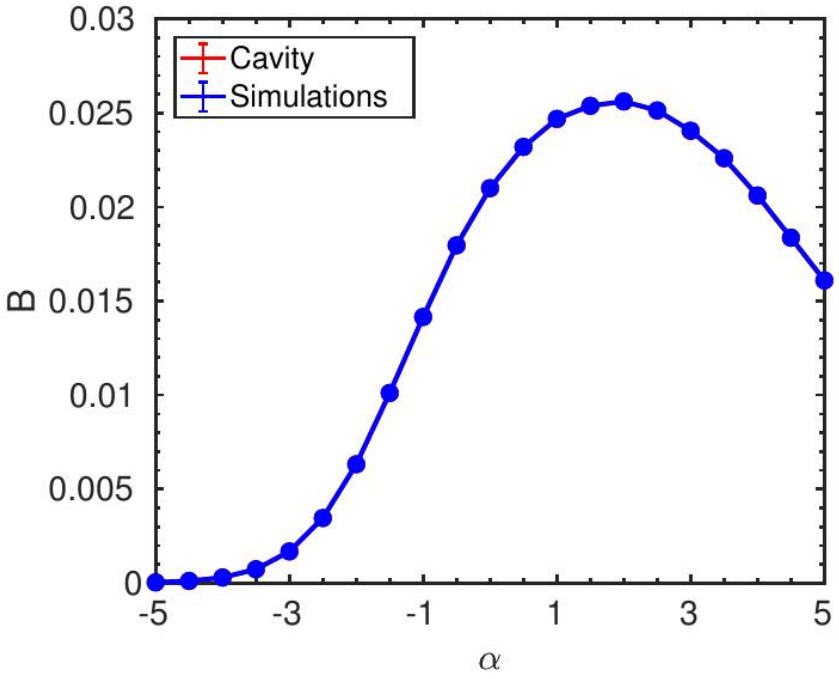
Simulations



Simulation results for exploration efficiency testing linearity $S(n) \sim Bn$ of a degree-biased random walker with $s(k) = k$ on the GC of an ER graph of size $N = 600$ with $c = 4$. Results are averaged over $N_s = 2000$ random graph realizations. The fraction of sites in the giant cluster is $\rho \simeq 0.98$. We estimate $B \simeq 0.7167 \pm 0.0002$. From [S. Pandey and RK, J Phys A 52, 085001 (2019)]

Simulations vs Single Instance Cavity

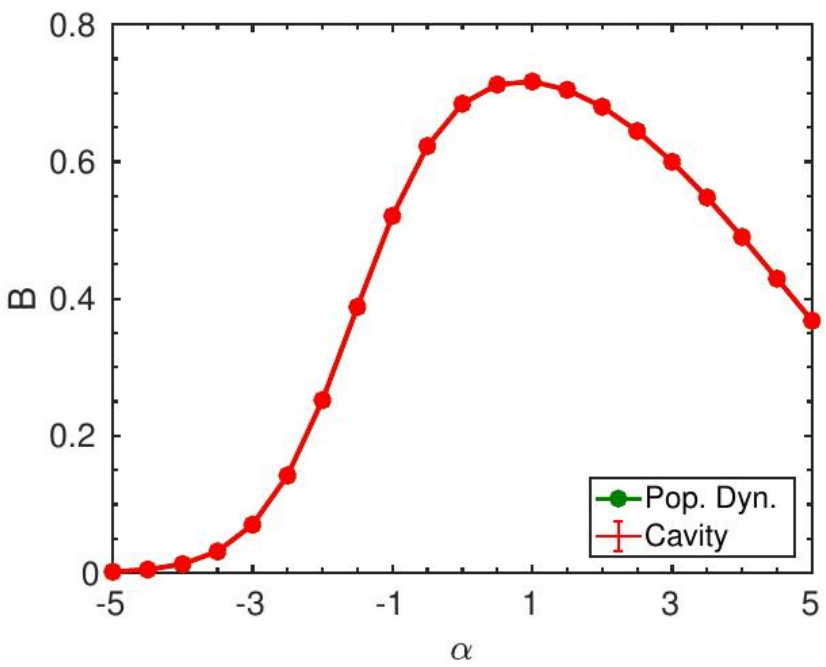


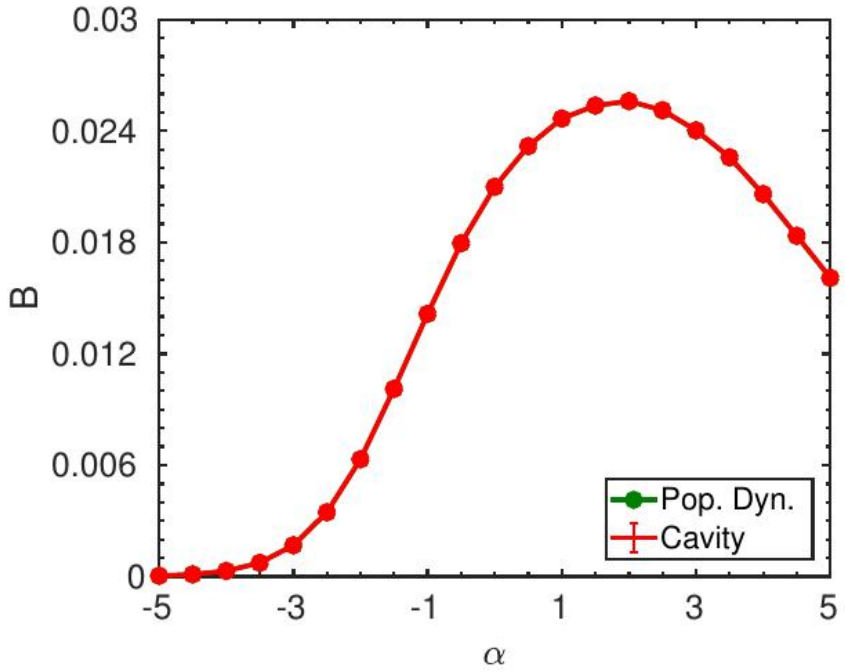


Comparing simulation and single instance cavity results. Left: exploration efficiency B of a degree-biased random walker with $s(k) = k^{\alpha}$ on the GC of an ER graph of size $N = 600$ with $c = 4$. Right: Search efficiency $B(\boldsymbol{\xi})$ of a power-law degree biased random walk computed for 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. Results are averaged over $N_s = 2000$ random graph realizations. Error bars are significantly smaller that the symbols.

From [S. Pandey and RK, J Phys A 52, 085001 (2019)]

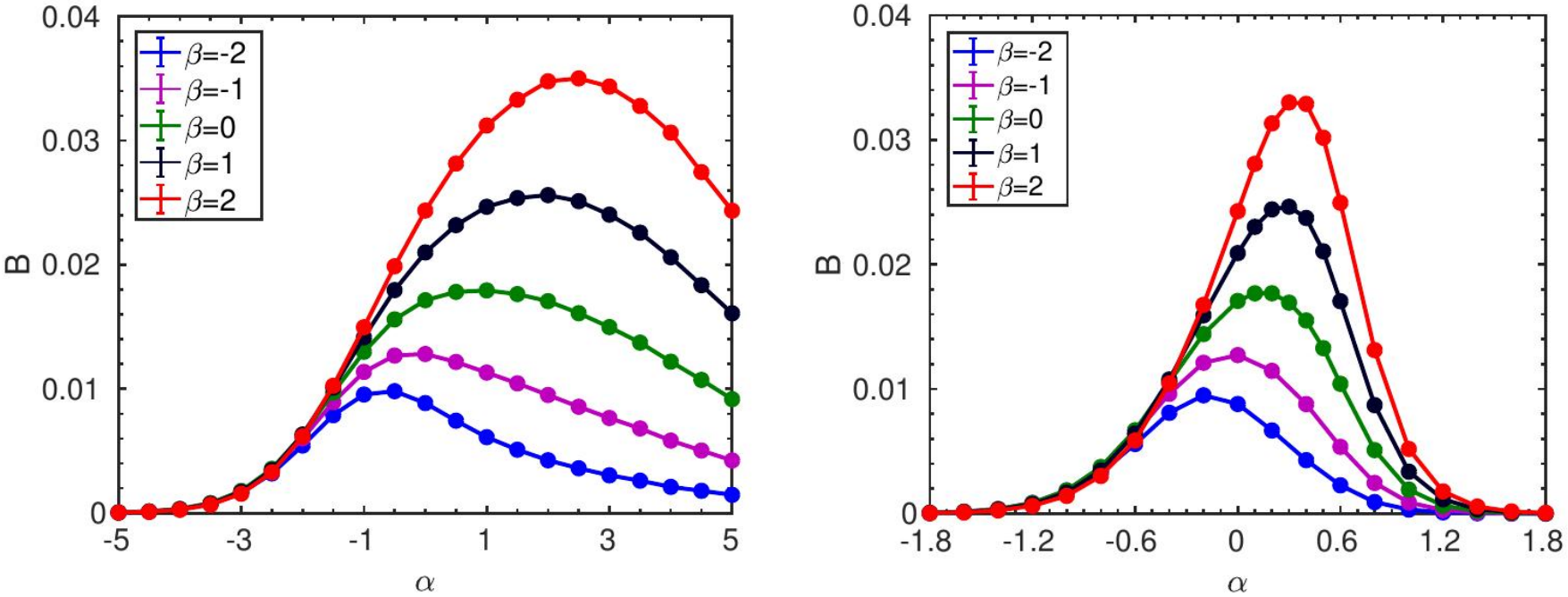
Population Dynamics vs Single Instance Cavity





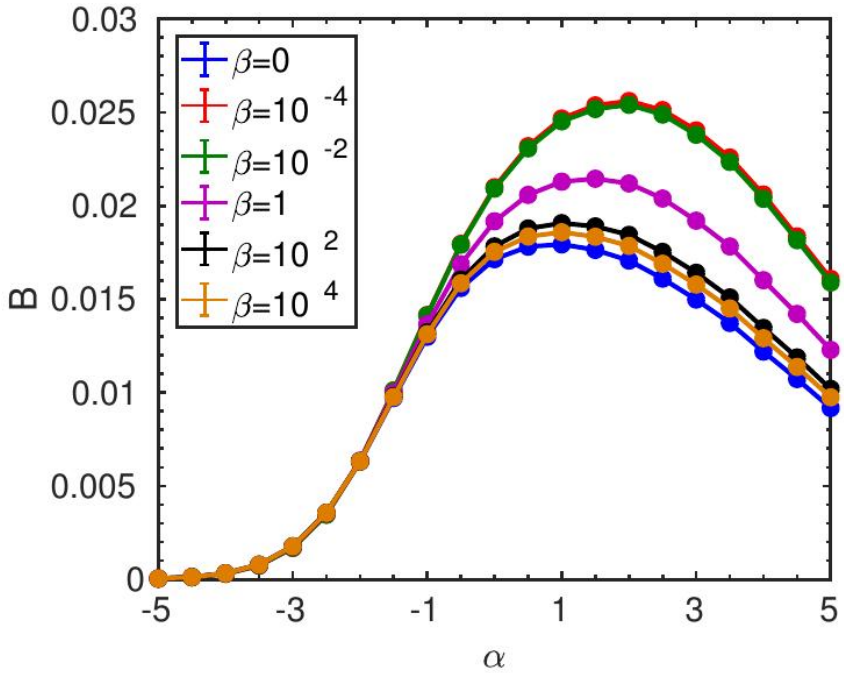
Comparison of single instance cavity and thermodynamic limit results. Left: exploration efficiency B of a degree-biased random walker with $s(k) = k^{\alpha}$ on the GC of an ER graph of size $N = 600$ with $c = 4$. Right: Search efficiency $B(\boldsymbol{\xi})$ of a power-law degree biased random walk 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. Results are averaged over $N_s = 2000$ random graph realizations. Error bars are significantly smaller that the symbols. From [S. Pandey and RK, J Phys A 52, 085001 (2019)]

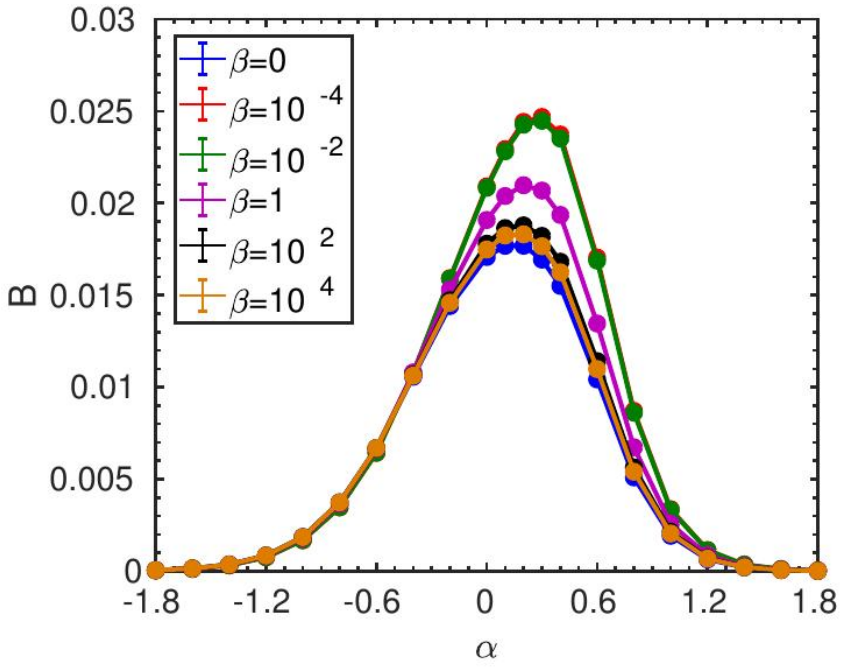
Hide and Seek



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 ER graphs with $c = 4$ and N=6000, averaged over $N_s = 2000$ instances. From [S. Pandey and RK, J Phys A 52, 085001 (2019)]

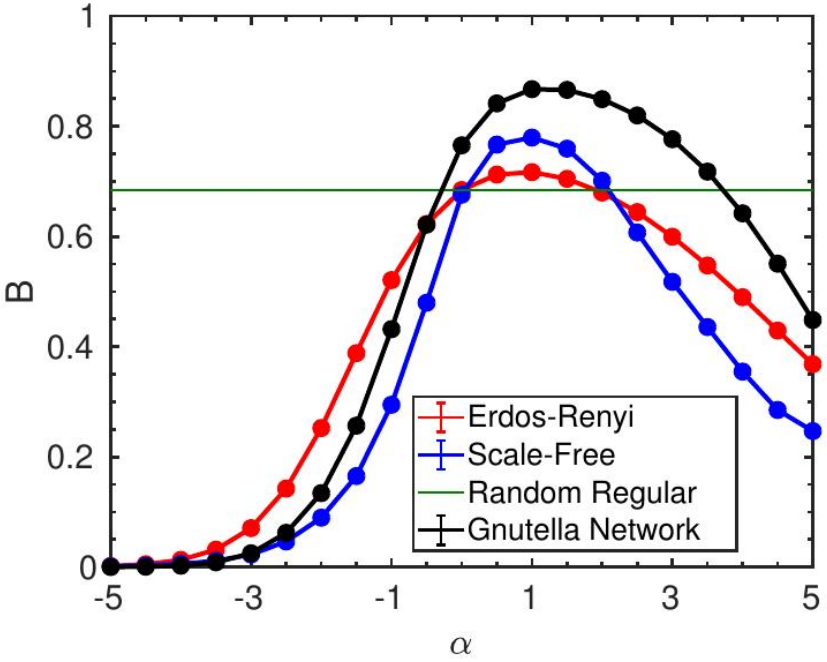
Hide and Seek





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 ER graphs with $c = 4$ and N=6000, averaged over $N_s = 2000$ instances. [S. Pandey and RK, J Phys A 52, 085001 (2019)]

Influence of Network Type

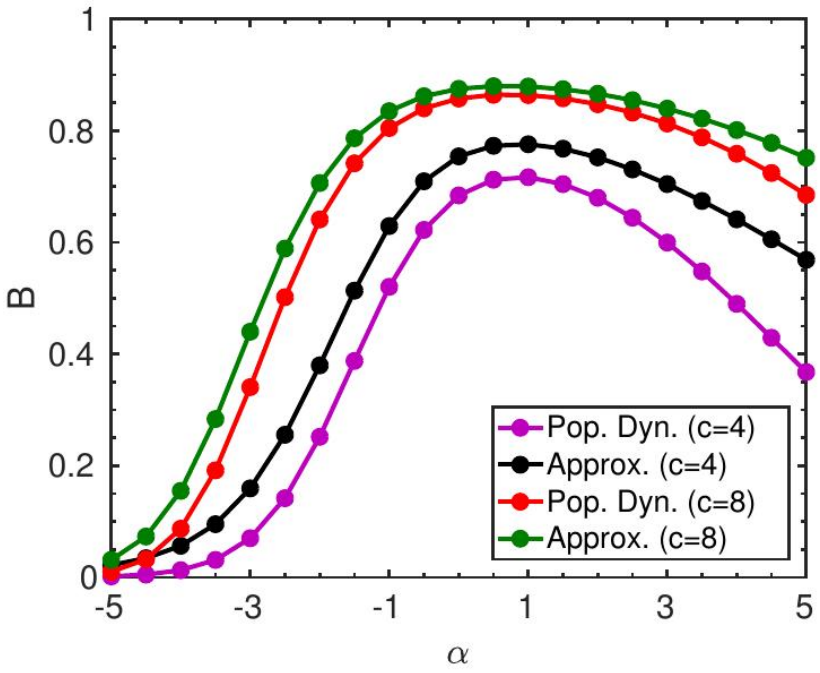


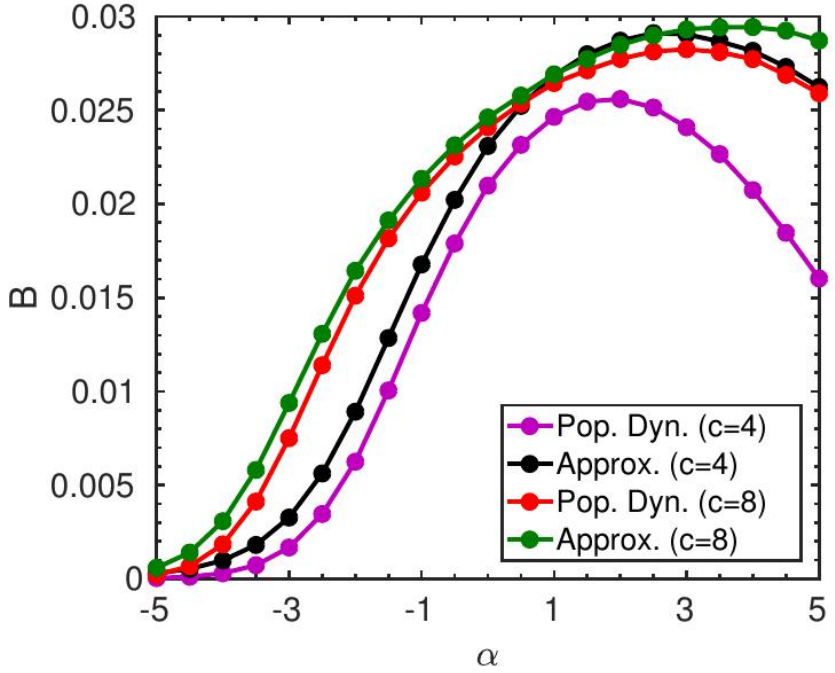
Comparison of network exploration efficiencies for four different graph types using the cavity method. Parameters are N=6000, and c=4 for ER and regular random graphs; for the scale-free graph we chose $\gamma = 2.65$, with $k_{\min} = 2$, $k_{\max} = 400$ giving a mean connectivity $c = 3.905$.

From[S. Pandey and RK, J Phys A 52, 085001 (2019)]

A Non-Backtracking Approximation

- Assuming that every non-backtracking step explores unseen parts of the network one can evaluate exploration and search efficiencies analytically.





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). Both panels compare results obtained via population dynamics for the thermodynamic limit with those of a non-backtracking approximation. From [S. Pandey and RK, J Phys A 52, 085001 (2019)]

Outline

Summary

- Evaluated network exploration and search efficiencies of degree biased random walkers on complex networks.
- Evaluation using cavity method for single large instances and in the thermodynamic limit.
- Thermodynamic limit required projection of results on giant cluster.
- Degree bias in search can increase efficiencies significantly.
- ...and exploit the heterogeneity of the network and/or hiding strategies
- Optimal search efficiencies $B(\xi) > \rho_h$ possible.
- Cavity and thermodynamic limit (population dynamics) results in excellent agreement with simulations.

- To do: (i) fluctuations, (ii) rare events ...

Summary

- Evaluated network exploration and search efficiencies of degree biased random walkers on complex networks.
- Evaluation using cavity method for single large instances and in the thermodynamic limit.
- Thermodynamic limit required projection of results on giant cluster.
- Degree bias in search can increase efficiencies significantly.
- ...and exploit the heterogeneity of the network and/or hiding strategies
- Optimal search efficiencies $B(\xi) > \rho_h$ possible.
- Cavity and thermodynamic limit (population dynamics) results in excellent agreement with simulations.

- To do: (i) fluctuations, (ii) rare events ...

Thank you!