Vacant sets and vacant nets:
Critical times for simple and modified random walks

Colin Cooper (KCL)

Bristol Algorithms Days 2016
Workshop on Efficient Algorithms and Lower Bounds

Joint work with Alan Frieze (CMU)
Context

- Discrete (distributed) processes on graphs
- Some particles (agents, messages, robots, viruses,..) move around the graph
- Movement between vertices can be random or deterministic
- Particles can be independent or interacting

- Purpose of process?
- Network exploration (search, estimate graph properties)
- Reach consensus of opinion (voting)
- Propagation of information (broadcasting, percolation, epidemics..)
Random walks and vacant sets

- Introduction: Random walks and Cover time
- Vacant set definition
- Typical results:
  - Vacant set in $G_{n,p}$, random $r$-regular graphs
- Vacant net definition
- Types of random walks
- Comparison of critical times
- Other results (Cover time)
Discrete random walk on a finite graph

- $G = (V, E)$ is a finite graph with $n$ vertices and $m$ edges. ($|V| = n, |E| = m$)
- Assume that $G$ is connected, so that all vertices in $G$ have at least one neighbour vertex,

Simple random walk: move to a randomly chosen neighbour at each step

Suppose that at step $t$ the walk is at vertex $X(t) = v$
Let $d(v) = \left| N(v) \right|$ be the degree of vertex $v$, then for $w \in N(v)$

$$\Pr(X(t + 1) = w) = \frac{1}{d(v)}$$
Cover time: Introduction

- $G = (V, E)$ is a connected graph. ($|V| = n, |E| = m$)

- Random walk $W_v$ on $G$ starting at $v \in V$
  Let $C_v$ be the expected time taken for $W_v$ to visit every vertex of $G$

- The **cover time** of $G$ is defined as $T_{cov} = \max_{v \in V} C_v$

- Starting vertex matters
- Graph with vertices $u, v, w$ $\quad C_v = C_u + 1$

\[ u \quad \underbrace{\quad \quad \quad \quad \quad \quad} \quad v \quad \underbrace{\quad \quad \quad \quad \quad \quad} \quad w \]
Brief history

- AKLLR\(^1\) For any connected graph \(T_{cov} \leq 4mn\)

- Application: Is there a path connecting vertex \(s\) to \(t\)?
  Test graph connectivity by random walk, in \(O(n^3)\) steps
  with \(O(\log n)\) storage.

- Good for exploring large networks.
  This led to increased interest in cover time

- Feige (1995). Bounds for any connected \(G\):
  \[
  (1 - o(1))n \log n \leq T_{cov} \leq (1 + o(1)) \frac{4}{27} n^3
  \]

\(^1\)Aleliunas, Karp, Lipton, Lovász and Rackoff. Random Walks, Universal Traversal Sequences, and the Complexity of Maze Problems. (1979)
Cover time $T_{\text{cov}}$ of random walk on graph $G$

$T_{\text{cov}}$ is the maximum expected time, over all start vertices $u$, for a random walk $\mathcal{W}_u$ to visit all vertices of $G$.

We found a method to estimate number of unvisited vertices at any step of the walk on random graphs and symmetric graphs.

Works well for cover time (no unvisited vertices)
Cover time $T_{cov}$ of random walk on graph $G$

$T_{cov}$ is the maximum expected time, over all start vertices $u$, for a random walk $W_u$ to visit all vertices of $G$.

We found a method to estimate number of unvisited vertices at any step of the walk on random graphs and symmetric graphs.

Works well for cover time (no unvisited vertices)

1. Erdös-Renyi random graphs $G_{n,p}$
   Let $np = c \log n$ and $(c - 1) \log n \to \infty$ then

   $$T_{cov} \sim c \log \left( \frac{c}{c - 1} \right) n \log n.$$

2. Random regular graphs $G_r$, where $3 \leq r = O(1)$ then

   $$T_{cov} \sim \frac{r - 1}{r - 2} n \log n.$$
Random walks and vacant sets

Purpose of random walk process?
Network exploration (search for info, estimate graph properties)

A short interlude for a demonstration
What is the component structure of vacant set?
Simple example
Graph and random walk

Graph of vacant set
Why might this be interesting?

Do we want to turn over every stone (cover time)
Or search the graph until no large chunks left (critical time)
Notation

Finite graph $G = (V, E)$.

$\mathcal{W}_u$ Simple random walk on $G$, starting at $u \in V$

The vacant set. Vertices not yet visited by the walk

Can think of vacant set $R(t)$ as coloured red, and visited vertices $B(t)$ as colored blue

$R(t)$ Set of vertices not visited by $\mathcal{W}_u$ up to time $t$

$\Gamma(t)$ Sub-graph of $G$ induced by vacant set $R(t)$

How large is $R(t)$?
What is the likely component structure of $\Gamma(t)$?
Random graphs: Giant component

If we increase the edge probability $p$ of $G_{n,p}$, at $p \sim 1/n$ usually a unique giant component emerges and continues to grow until the graph is connected.
Similarly we can start with the empty graph on $n$ vertices and insert the edges of $K_n$ in some random order (graph process). The giant emerges when about $n/2$ edges have been added.
Evolution of vacant set graph $\Gamma(t)$

It is a sort of random graph process in reverse
Initially all vertices and edges are unvisited
As the walk progresses we visit new vertices and the vacant set $\Gamma(t)$ is reduced from the whole graph $G$ to a graph with no vertices

In the context of sparse random graphs, as the unvisited vertex set $\mathcal{R}(t)$ gets smaller, the edges inside $\Gamma(t)$ will get sparser and sparser.

Small sets of vertices don’t induce many edges
One might expect that at some time $\Gamma(t)$ will break up into small components
We say that $\Gamma(t)$ is sub-critical at step $t$, if all of its components are of size $O(\log n)$

We say that $\Gamma(t)$ is super-critical at step $t$, if it has a unique giant component, (of size $\Theta(R(t))$) and all other components are of size $O(\log n)$

In the cases we consider there is a $t^*$, which is a (whp) threshold for transition from super-criticality to sub-criticality
Vacant set of $G_{n,p}$

We assume that

$$p = \frac{c \log n}{n}$$

where $(c - 1) \log n \to \infty$ with $n$, and $c = n^{o(1)}$. Let

$$t(\epsilon) = n \left( \log \log n + (1 + \epsilon) \log c \right)$$

**Theorem**

*Let $\epsilon > 0$ be a small constant

Then *whp* we have

(i) $\Gamma(t)$ is super-critical for $t \leq t(-\epsilon)$

(ii) $\Gamma(t)$ is sub-critical for $t \geq t(\epsilon)$

Giant component of $\mathcal{R}(t)$ until $t > n \log \log n$

For $c > 1$ constant, Cover time $T_{cov}$ of $G_{n,p}$ is $T_{cov} \sim n \log n$
Random graphs $G_{n,r}$

For $r \geq 3$, constant, let

$$t^* = \frac{r(r - 1) \log(r - 1)}{(r - 2)^2} n$$

**Theorem**

Let $\epsilon > 0$ be a small constant. Then whp we have

(i) $\Gamma(t)$ is super-critical for $t \leq (1 - \epsilon)t^*$

(ii) For $t \leq (1 - \epsilon)t^*$, size of giant component is $\Omega(n)$

(iii) $\Gamma(t)$ is sub-critical for $t \geq (1 + \epsilon)t^*$

e.g. for 3-regular random graphs $r = 3$, and $t^* = (6 \log 2) n$

Giant component for about $t^* = (6 \log 2)n$ steps

Cover time $T_{cov} \sim 2n \log n$
Related Work

Benjamini and Sznitman; Windisch:
Considered the $d$-dimensional toroidal grids $d \geq 5$.
Super-critical below $C_1 n$, sub-critical above $C_2 n$

Černy, Teixeira and Windisch:
Considered random $r$-regular graphs $G_{n,r}$
They showed sub-criticality for $t \geq (1 + \epsilon)t^*$
and existence of a unique giant component for $t \leq (1 - \epsilon)t^*$
These proofs use the concept of random interlacements of continuous time random walks
Our proof: Discrete time

- Simple. Based on established random graph results
- Gives results for $G_{n,p}$
- Completely characterizes the component structure
- Proves that in the super-critical phase $t \leq t^*$, the second largest component of $G_{n,r}$ has size $O(\log n)$ whp
  Gives the small tree structure of $\Gamma(t)$

Subsequent Work: Černy, Teixeira and Windisch:
Consider random $r$-regular graphs $G_{n,r}$
Investigate scaling window around $t^*$ using annealed model
How to estimate size of vacant set (I)

- Estimate the un-visit probability of states
  The probability a given state has has not been visited after $t$ steps of the process

- Related quantity
  The probability a given state has has not been visited after $t$ steps of the process, starting from stationarity

- The expected hitting time of state $v$ from stationarity can be approximated by

$$E_{\pi}H_v \sim R_v / \pi_v$$

where $R_v$ is expected number of returns to $v$ during a suitable mixing time

- (Informally) Waiting time of first visit to $v$ tends to geometric distn, success probability $p_v \sim \pi_v / R_v$
How to estimate size of vacant set (I)

- Estimate the un-visit probability of states
  The probability a given state has has not been visited after $t$ steps of the process

- Related quantity
  The probability a given state has has not been visited after $t$ steps of the process, starting from stationarity

\[
E^\pi H_v \sim \frac{R_v}{\pi_v}
\]

(Informally) Waiting time of first visit to $v$ tends to geometric distn, success probability $p_v \sim \frac{\pi_v}{R_v}$
How to estimate size of vacant set (I)

- Estimate the **un-visit probability** of states
  The probability a given state has has not been visited after $t$ steps of the process

- Related quantity
  The probability a given state has has not been visited after $t$ steps of the process, starting from stationarity

- The expected hitting time of state $v$ from stationarity can be approximated by

$$E_{\pi} H_v \sim R_v / \pi_v$$

where $R_v$ is expected number of returns to $v$ during a suitable mixing time

- (Informally) Waiting time of first visit to $v$ tends to geometric distn, success probability $p_v \sim \pi_v / R_v$
How to estimate size of vacant set (II)

Size of vacant set $\mathcal{R}(t)$?
For graphs which are (reasonably) rapidly mixing

$$E(|\mathcal{R}(t)|) \sim \sum_v e^{-t\pi_v/R_v}$$

$\pi_v$ is stationary probability of vertex $v$, i.e. $\pi_v = d(v)/2m$

For a random walk starting out from $v$, $R_v$ is the expected number of returns to $v$ during the mixing time. $R_v$ depends only on the structure of the graph around $v$
Component structure of vacant set of random $r$-regular graphs for $r \geq 3$, constant.
Threshold

Let

\[ t^* = \frac{r(r - 1) \log(r - 1)}{(r - 2)^2} n. \]

Theorem

Let \( \epsilon > 0 \) be a small constant. Then \textbf{whp} we have

(i) \( \Gamma(t) \) is super-critical for \( t \leq (1 - \epsilon)t^* \),

(ii) For \( t \leq (1 - \epsilon)t^* \), size of giant component is \( \Omega(n) \)

(iii) \( \Gamma(t) \) is sub-critical for \( t \geq (1 + \epsilon)t^* \) and
Proof outline for $r$-regular random graph

- Generate the graph in the configuration model using the random walk
- Graph $\Gamma(t)$ induced by vacant set $R(t)$ is random
- Estimate un-visit probability of vertices to find size of $R(t)$
- Estimate degree sequence $d$ of $\Gamma(t)$ in the configuration model, using size of vacant set $R(t)$, and number of unvisited edges $U(t)$
- Given the degree sequence $d$ of $\Gamma(t)$, we can use Molloy-Reed condition for existence of giant component in a random graph with fixed degree sequence
- Estimate number of small trees in configuration model
The graph induced by the vacant set is random

**Lemma**

Consider a random walk on $G_r$. Conditional on $N = |\mathcal{R}(t)|$ and degree sequence $d = d_{\Gamma(t)}(v)$, $v \in \mathcal{R}(t)$, then $\Gamma(t)$ is distributed as $G_{N,d}$, the random graph with vertex set $[N]$ and degree sequence $d$.

**Proof**  
Basic idea: Reveal $G_r$ using the random walk. Suppose that we condition on $\mathcal{R}(t)$ and the *history of the walk*, $\mathcal{H} = (W_u(0), W_u(1), \ldots, W_u(t))$. If $G_1, G_2$ are graphs with vertex set $\mathcal{R}(t)$ and if they have the same degree sequence then substituting $G_2$ for $G_1$ will not conflict with $\mathcal{H}$. Every extension of $G_1$ is an extension of $G_2$ and vice-versa. □

Thus we only need:
Good model of component structure of $G_{N,d}$
High probability estimates of the degree sequence $D_s(t)$ of $\Gamma(t)$. 
Degree sequence of $\Gamma(t)$

**Vacant set size**

$|\mathcal{R}(t)| = (1 + o(1))N_t$ where $N_t = ne^{-(r-2)t \over (r-1)n}$

**Vertex degree**

Let $D_s(t)$ the number of unvisited vertices of $\Gamma(t)$ with $r - s$ visited neighbours and of degree $s$ in $\Gamma(t)$

For $0 \leq s \leq r$, and for ranges of $t$ given below, **whp**

$$D_s(t) \sim N_t \binom{r}{s} p_t^s (1 - p_t)^{r-s}$$

where

$$p_t = e^{-\frac{(r-2)^2 t}{(r-1)r n}}$$
Molloy-Reed Condition

Theorem

Let $\lambda_0, \lambda_1, \ldots, \lambda_r \in [0, 1]$ be such that $\lambda_0 + \lambda_1 + \cdots + \lambda_r = 1$. Suppose that $d = (d_1, d_2, \ldots, d_N)$ satisfies 
\[ |\{j : d_j = s\}| = (1 + o(1))\lambda_s N \quad \text{for } s = 0, 1, \ldots, r. \]

Let $G_{n,d}$ be chosen randomly from graphs with vertex set $[N]$ and degree sequence $d$. Let

\[ L = \sum_{s=0}^{r} s(s - 2)\lambda_s. \]

(a) If $L < 0$ then \textbf{whp} $G_{n,d}$ is sub-critical.

(b) If $L > 0$ then \textbf{whp} $G_{n,d}$ is super-critical.

Furthermore the unique giant component has size $\beta n$ where $\beta$ is the solution to an equation derived from the degree sequence.
Threshold for collapse of giant component

Degree sequence of $\Gamma(t)$ is (approximately) binomial $Bin(r, p_t)$
where $p_t = e^{-\frac{t(r-2)^2}{n(r-1)r}}$

Once we know the degree sequence we can use the Molloy-Reed criterion to see whether or not there is a giant component. $G$ has a giant component iff $L > 0$, where

$$L = \sum_v d_v(d_v - 2).$$

Direct calculation gives $t^* = \frac{r(r-1)\log(r-1)}{(r-2)^2} n$ as the critical value
Heuristically, \( t^* = \frac{r(r-1)\log(r-1)}{(r-2)^2} n \) can be obtained from the degree sequence of unvisited vertices.

Branching outward from an unvisited vertex, the probability an edge goes to another unvisited vertex:

\[
p_t = e^{-\frac{(r-2)^2 t}{(r-1)n}}
\]

We need branching factor \((r - 1)p_t \geq 1\), to have a chance to get a large component.

At \( t^* = \frac{r(r-1)\log(r-1)}{(r-2)^2} n \)

\[
(r - 1)p_t = (r - 1)e^{-\frac{(r-2)^2 t}{(r-1)n}}
= (r - 1)e^{-\log(r-1)}
= 1
\]
Example: $r = 3$. Vacant set as a function of $\tau = t/n$

Proportion of vertices in vacant set $N(t)/n \sim e^{-t/n((r-2)/(r-1))}$

Proportion of vertices in unvisited tree components
Threshold: \( r = 3, \ t^* = 6 \log 2 \)

\[
t^* = \frac{r(r - 1) \log(r - 1)}{(r - 2)^2} n
\]

Proportion of vertices in vacant set, and on small tree components

The cusp is at \( t^* = 6 \log 2 \approx 4.16 \), with \( \mu^* = 1/8 \).
Vacant Sets and Vacant Nets
Vacant sets and vacant nets

- What to consider? Unvisited vertices or unvisited edges?
- What is the difference? Vacant net will be more connected
- Even if the vacant set is small, the vacant net can be huge
- Vertex cover time or Edge cover time
- Why interesting?
  How long does a random walk take to break up large networks into small chunks
  Once the giant component has collapsed, most components are constant size.
  We can be sure the walk been near to all vertices, even if not actually visited
- Interesting and challenging to analyse
Vacant set or vacant net?
Threshold comparison for r-regular graphs

Thresholds $\sim cn$ where $c$ is the following constant

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Simple random walk</th>
<th>Example: $r = 4$</th>
<th>$r = 100$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vacant set</td>
<td>$\frac{r(r-1)}{(r-2)^2} \log(r - 1)$</td>
<td>$3 \log 3 \approx 3.29$</td>
<td>4.7</td>
</tr>
<tr>
<td>Vacant net</td>
<td>$\frac{r}{2} \times \frac{r^2-2r+2}{(r-2)^2} \log(r - 1)$</td>
<td>$5 \log 3 \approx 5.49$</td>
<td>235</td>
</tr>
</tbody>
</table>

As $r \to \infty$ the constants $c$ tend to $\log r$ and $(r/2) \log r$

It takes $r/2$ longer to break up the vacant net (much longer)
Can we speed up the collapse of the giant?

We checked the following types of random walk:

- Simple random walk (as above)
- Non-backtracking random walk
- Random walk which prefers unvisited edges
  (If unvisited edge(s) at current vertex choose one at random, otherwise choose random neighbour)
Compare random walks

- Simple random walk
- Non-backtracking random walk
- Random walk which prefers unvisited edges

For any reversible random walk the expected time to visit all vertices is at least $\Theta(n \log n)$

Reversible? Transition probabilities proportional to positive edge weights on the graph

Theoretical lower bound. Any walk takes $n$ steps to visit $n$ vertices

Any walk with cover time $o(n \log n)$ must be non-reversible

Various suggestions: Non-backtrack, prefer unvisited vertices or edges,...
Asymptotic critical times

<table>
<thead>
<tr>
<th></th>
<th>Vacant set</th>
<th>Simple random walk</th>
<th>Non-backtrack random walk</th>
<th>Unvisited edge process, ((r \text{ even}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>V. Set</td>
<td>(n \log r)</td>
<td>(n \log r)</td>
<td>(n \log r)</td>
<td>(n \log r)</td>
</tr>
<tr>
<td>V. Net</td>
<td>(n \frac{r}{2} \log r)</td>
<td>(n \frac{r}{2} \log r)</td>
<td>(n \frac{r}{2} \log r)</td>
<td>(n \frac{r}{2})</td>
</tr>
</tbody>
</table>

All vacant set thresholds tend to \(n \log r\): No speed up.
Unvisited edge process speeds up collapse of vacant net by a factor of \(\log r\)

Actual constants: 4-regular graphs

<table>
<thead>
<tr>
<th>Degree</th>
<th>Simple random walk</th>
<th>Non-backtrack random walk</th>
<th>Unvisited edge process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vacant set</td>
<td>3 (\log 3 \approx 3.29)</td>
<td>2 (\log 3 \approx 2.20)</td>
<td>4/3</td>
</tr>
<tr>
<td>Vacant net</td>
<td>5 (\log 3 \approx 5.49)</td>
<td>3 (\log 3 \approx 3.29)</td>
<td>2</td>
</tr>
</tbody>
</table>
Cover time of unvisited edge process

Tom Friedetzky’s simulation results
For $r$ even, cover time $= rn/2$? ($r = d$ in figure)
Vertex cover time: Comparison

**Theorem**

Let $T_{cov}^V(G)$ denote the vertex cover time of the associated walk. The following asymptotic results hold w.h.p.

<table>
<thead>
<tr>
<th>Type</th>
<th>Simple random walk</th>
<th>Non-backtrack random walk</th>
<th>Unvisited edge process</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{cov}^V(G)$</td>
<td>$\frac{r-1}{r-2} n \log n$</td>
<td>$n \log n$</td>
<td>$\frac{r}{2} n$</td>
</tr>
</tbody>
</table>

$r$ even
Edge process: Vertex cover time for $r$ even

- For $r$ constant: The vacant net collapses at the vertex cover time $rn/2$

- As $r \to \infty$, the cover time tends to $n \log n$ from below

- To prove cover time $rn/2$ was ok for $r \geq 6$
  For $r = 4$ consider in detail the (de)-evolution of the giant component for the unvisited edge process. Cannot use Molloy-Reed condition.

- For $r$ odd, not so much known or understood. Can only prove cover time $O(n \log n)$. (See figure)
Closing observations

- Random walks on random graphs ($G(n, p)$) and random $r$-regular graphs exhibit threshold behavior. The vacant set collapses rapidly at the threshold.
- The size of the giant component can be estimated in the super-critical range.
- The number of small tree components of a given size can be estimated.

- The technique can be applied to other problems:
  - Vacant net: sub-graph induced by the unvisited edges.
  - Upper bounds on sub-critical threshold for hypercube, high degree grids.
  - We can compare thresholds for different types of random walks.
THANK YOU

QUESTIONS
Advert for RA
Advert for RA at PostDoc Level

- For: 18 months, starting summer 2016
- Based at: King’s College London
- Topic: Random walks and random processes
- Research: Mixture of theory and experimental evaluation
- Contact: colin.cooper@kcl.ac.uk, or ask
- Scope to be creative