A Transport Redundancy Approach to Spatial Clustering

Mingyue Gao

Wake Forest University
gaom212@wfu.edu

May 2, 2014

Advisor: Kenneth S. Berenhaut, Ph.D.

The Interest

In this thesis, we consider a new random walks-based method for measuring clustering on finite graphs. The method is versatile and can provide valuable insight into the locations of centers of clusters and outliers. Some results and applications are discussed.

The Problem

The Problem

- Consider a random walk on a connected graph $G = G(V, E)$ with N nodes, i.e. $|V| = N$.

The Problem

- Consider a random walk on a connected graph $G = G(V, E)$ with N nodes, i.e. $|V| = N$.
- Among the N nodes, all elements of the set $\mathcal{S} = \{s_1, s_2, ..., s_n\} \subseteq V$ with $|\mathcal{S}| = n < N$ are marked as "maps". One of the maps is demarcated as the "target", denoted as $\dagger$.

The Problem

- Consider a random walk on a connected graph $G = G(V, E)$ with N nodes, i.e. $|V| = N$.
- Among the N nodes, all elements of the set $\mathcal{S} = \{s_1, s_2, ..., s_n\} \subseteq V$ with $|\mathcal{S}| = n < N$ are marked as "maps". One of the maps is demarcated as the "target", denoted as $\dagger$.
- In determining its next step, the walker chooses uniformly randomly among its neighbors, until such a point that a map is landed on.

The Problem

- Consider a random walk on a connected graph $G = G(V, E)$ with N nodes, i.e. $|V| = N$.
- Among the N nodes, all elements of the set $\mathcal{S} = \{s_1, s_2, \ldots, s_n\} \subseteq V$ with $|\mathcal{S}| = n < N$ are marked as "maps". One of the maps is demarcated as the "target", denoted as $\dagger$.
- In determining its next step, the walker chooses uniformly randomly among its neighbors, until such a point that a map is landed on.
- At that point, the walker is sent on a shortest path to the target.

The Problem

- Consider a random walk on a connected graph $G = G(V, E)$ with N nodes, i.e. $|V| = N$.
- Among the N nodes, all elements of the set $\mathcal{S} = \{s_1, s_2, ..., s_n\} \subseteq V$ with $|\mathcal{S}| = n < N$ are marked as "maps". One of the maps is demarcated as the "target", denoted as $\dagger$.
- In determining its next step, the walker chooses uniformly randomly among its neighbors, until such a point that a map is landed on.
- At that point, the walker is sent on a shortest path to the target.
- Of particular interest will be the expected amount of time that a randomly placed walker takes to reach a randomly chosen target among the maps.

Notations

Notations

Let $\{X_0, X_1, \ldots, X_{T_0}, \ldots\}$ be a random walk starting at $X_0 = x_0$.

Notations

Let $\{X_0, X_1, \ldots, X_{T_0}, \ldots\}$ be a random walk starting at $X_0 = x_0$. We then have

$$T_0 \overset{def}{=} \min\{i \geq 0 | X_i \in \mathcal{S}\},$$

Notations

Let $\{X_0, X_1, \ldots, X_{T_0}, \ldots\}$ be a random walk starting at $X_0 = x_0$. We then have

$$
\begin{aligned}
T_0 &\overset{def}{=} \min\{i \geq 0 | X_i \in \mathcal{S}\}, \\
T_1 &\overset{def}{=} dist(X_{T_0}, \dagger) \qquad\qquad \text{and}
\end{aligned}
$$

Notations

Let $\{X_0, X_1, \ldots, X_{T_0}, \ldots\}$ be a random walk starting at $X_0 = x_0$. We then have

$$\begin{aligned} T_0 &\stackrel{def}{=} \min\{i \geq 0 | X_i \in \mathcal{S}\}, \\ T_1 &\stackrel{def}{=} dist(X_{T_0}, \dagger) \qquad\qquad \text{and} \\ T &\stackrel{def}{=} T_0 + tT_1, \end{aligned}$$

Notations

Let $\{X_0, X_1, \ldots, X_{T_0}, \ldots\}$ be a random walk starting at $X_0 = x_0$. We then have

$$\begin{aligned} T_0 &\stackrel{def}{=} \min\{i \geq 0 | X_i \in \mathcal{S}\}, \\ T_1 &\stackrel{def}{=} dist(X_{T_0}, \dagger) \qquad \text{and} \\ T &\stackrel{def}{=} T_0 + tT_1, \end{aligned}$$

where $t \geq 0$ is a delay factor.

Notations

Let

$$e(\mathcal{S}, \dagger, X_0) \stackrel{def}{=} E[T|\mathcal{S}, \dagger, X_0]$$

denote the expected time to hit $\dagger \in \mathcal{S}$, beginning at $X_0 \in V$, and

Notations

Let

$$e(\mathcal{S}, \dagger, X_0) \stackrel{def}{=} E[T|\mathcal{S}, \dagger, X_0]$$

denote the expected time to hit $\dagger \in \mathcal{S}$, beginning at $X_0 \in V$, and

$$\bar{e}(\mathcal{S}, \dagger) \stackrel{def}{=} E[e(\mathcal{S}, \dagger, X_0)]$$

be the expected time to reach $\dagger$ from a random starting point $X_0 \in V$.

Notations

Let

$$e(\mathcal{S}, \dagger, X_0) \stackrel{def}{=} E[T|\mathcal{S}, \dagger, X_0]$$

denote the expected time to hit $\dagger \in \mathcal{S}$, beginning at $X_0 \in V$, and

$$\bar{e}(\mathcal{S}, \dagger) \stackrel{def}{=} E[e(\mathcal{S}, \dagger, X_0)]$$

be the expected time to reach $\dagger$ from a random starting point $X_0 \in V$.

If X_0 is taken uniformly randomly from V, we then have

$$\bar{e}(\mathcal{S}, \dagger) = \frac{1}{N}\left(\sum_{v \in V} e(\mathcal{S}, \dagger, v)\right).$$

Notations

Definition

Define $\mathcal{E}$, via

$$\mathcal{E}(\mathcal{S}) \stackrel{def}{=} \frac{1}{n}\left(\sum_{s_i \in \mathcal{S}} \bar{e}(\mathcal{S}, s_i)\right).$$

Notations

Definition

Define $\mathcal{E}$, via

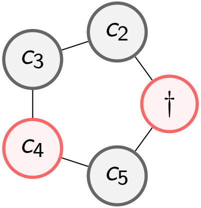
$$\mathcal{E}(\mathcal{S}) \stackrel{def}{=} \frac{1}{n}\left(\sum_{s_i \in \mathcal{S}} \bar{e}(\mathcal{S}, s_i)\right).$$

An important quantity, in what follows, $\mathcal{E}(\mathcal{S})$, is the expected amount of time to reach a uniformly randomly chosen target (from $\mathcal{S}$) beginning at a uniformly randomly chosen starting point.

A Simple One-Dimensional Cycle Example

A Simple One-Dimensional Cycle Example

Consider a one-dimensional cycle $\mathcal{C}_5$ and $\mathcal{S} = \{c_1, c_4\}$ is the set of "maps". Let $t = 1$. First, set c_1 to be the "target". Let e_i denote $e(\mathcal{S}, \dagger, c_i)$.



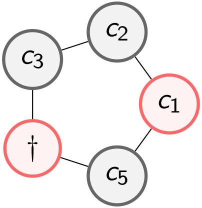
We then have $e_1 = 0$, $e_2 = 1 + (e_1 + e_3)/2$, $e_3 = 1 + (e_2 + e_4)/2$, $e_4 = 2$ and $e_5 = 1 + (e_4 + e_1)/2$.

We then have $e_1 = 0$, $e_2 = 1 + (e_1 + e_3)/2$, $e_3 = 1 + (e_2 + e_4)/2$, $e_4 = 2$ and $e_5 = 1 + (e_4 + e_1)/2$.
In matrix form, this gives $A\mathbf{e} = b$ where

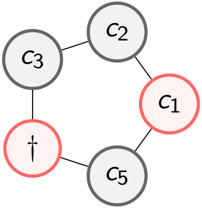
$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & 0 \\ 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 & 0 \\ -\frac{1}{2} & 0 & 0 & -\frac{1}{2} & 1 \end{bmatrix}_{5\times 5} \quad \text{and} \quad b = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 2 \\ 1 \end{bmatrix}$$

$$\mathbf{e} = A^{-1}\mathbf{b} = \begin{bmatrix} 0 \\ 8/3 \\ 10/3 \\ 2 \\ 2 \end{bmatrix} \quad \text{and} \quad \bar{e}(\mathcal{S}, c_1) = \frac{1}{5}\left(\sum_{c_i \in \mathcal{C}_5} e_i\right) = 2$$

Set c_4 to be the target.

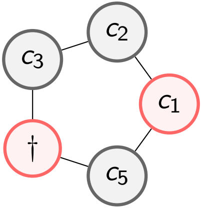


Set c_4 to be the target.



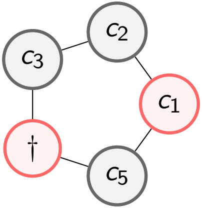
Similar calculation gives $\bar{e}(\mathcal{S}, c_4) = 2$.

Set c_4 to be the target.



Similar calculation gives $\bar{e}(\mathcal{S}, c_4) = 2$.
Recall $\bar{e}(\mathcal{S}, c_1) = 2$.

Set c_4 to be the target.



Similar calculation gives $\bar{e}(\mathcal{S}, c_4) = 2$.
Recall $\bar{e}(\mathcal{S}, c_1) = 2$.

$$\mathcal{E}(\mathcal{S}) = \frac{1}{2}\left(\sum_{s_i \in \mathcal{S}} \bar{e}(\mathcal{S}, s_i)\right) = 2$$

$\square$

- Note that A is diagonally dominant, sparse and nearly symmetric (with unit diagonal).

- Note that A is diagonally dominant, sparse and nearly symmetric (with unit diagonal).
- If $c_i \in \mathcal{S}$, then the ith row of A has only one non-zero entry.

- Note that A is diagonally dominant, sparse and nearly symmetric (with unit diagonal).
- If $c_i \in \mathcal{S}$, then the ith row of A has only one non-zero entry.
- If $c_i \notin \mathcal{S}$, then the ith row of A has two more $-1/2$ entries corresponding to the neighboring c_{i+1} and c_{i-1} on the cycle.

- Note that A is diagonally dominant, sparse and nearly symmetric (with unit diagonal).
- If $c_i \in \mathcal{S}$, then the ith row of A has only one non-zero entry.
- If $c_i \notin \mathcal{S}$, then the ith row of A has two more $-1/2$ entries corresponding to the neighboring c_{i+1} and c_{i-1} on the cycle.
- The matrix becomes increasingly sparse, for a fixed graph, as the number of maps increases.

Preliminaries and Notations for $\mathcal{C}_N$

Preliminaries and Notations for $\mathcal{C}_N$

For convenience, for $s_1, s_2 \in \mathcal{S}$, where $s_1 < s_2$, set

$$\mathcal{I} = \mathcal{I}_{s_1,s_2} \stackrel{def}{=} (s_1, s_2) = \{s_1 + 1, s_1 + 2, \ldots, s_2 - 1\},$$

Preliminaries and Notations for $\mathcal{C}_N$

For convenience, for $s_1, s_2 \in \mathcal{S}$, where $s_1 < s_2$, set

$$\mathcal{I} = \mathcal{I}_{s_1,s_2} \overset{def}{=} (s_1, s_2) = \{s_1 + 1, s_1 + 2, \ldots, s_2 - 1\},$$

$$a \overset{def}{=} dist(s_1, \dagger) \quad \text{and} \quad b \overset{def}{=} dist(s_2, \dagger),$$

Preliminaries and Notations for $\mathcal{C}_N$

For convenience, for $s_1, s_2 \in \mathcal{S}$, where $s_1 < s_2$, set

$$\mathcal{I} = \mathcal{I}_{s_1,s_2} \stackrel{def}{=} (s_1, s_2) = \{s_1 + 1, s_1 + 2, \ldots, s_2 - 1\},$$

$$a \stackrel{def}{=} dist(s_1, \dagger) \quad \text{and} \quad b \stackrel{def}{=} dist(s_2, \dagger),$$

and

$$K_t \stackrel{def}{=} \frac{1}{|\mathcal{I}|} \sum_{v_i \in \mathcal{I}} e\left(\mathcal{S}, \dagger, v_i\right).$$

Existing Work for $\mathcal{C}_N$

Lemma

(Wei) Suppose $s_1, s_2 \in \mathcal{S}$, $s_1 < s_2$ and $\mathcal{I} \cap \mathcal{S} = \emptyset$. We then have

$$K_1 = \frac{d^2 + d}{6} + \frac{a + b}{2},$$

where

$$d = |\mathcal{I}| + 1.$$

Existing Work for $\mathcal{C}_N$

Lemma

(Wei) Suppose $s_1, s^*, s_2 \in \mathcal{S}$, $s_1 < s^* < s_2$, *and* $\mathcal{I} \cap \mathcal{S} = s^*$, *and* $r = dist(s^*, s_1)$*. Then,* K_1 *thought of as a function of* r*, is minimized when* $r = \dfrac{d-1}{2} + \dfrac{b-a}{2d}$*, where* $d = |\mathcal{I}| + 1$*.*

Existing Work for $\mathcal{C}_N$

Definition

Suppose on a cycle $\mathcal{C}_N$, $\mathcal{S} = \{s_1, s_2, \ldots, s_n\}$, with $s_1 < \ldots < s_n$. The set $\mathcal{S}$ is said to be "evenly distributed" on $\mathcal{C}_N$, if for $\mathcal{T} = \{s_2 - s_1, s_3 - s_2, \ldots, s_n - s_{n-1}, s_1 - s_n + N\}$, we have $\max(\mathcal{T}) - \min(\mathcal{T}) \leq 1$.

Existing Work for $\mathcal{C}_N$

Definition

Suppose on a cycle $\mathcal{C}_N$, $\mathcal{S} = \{s_1, s_2, \ldots, s_n\}$, with $s_1 < \ldots < s_n$. The set $\mathcal{S}$ is said to be "evenly distributed" on $\mathcal{C}_N$, if for $\mathcal{T} = \{s_2 - s_1, s_3 - s_2, \ldots, s_n - s_{n-1}, s_1 - s_n + N\}$, we have $\max(\mathcal{T}) - \min(\mathcal{T}) \leq 1$.

Theorem

(Wei) The quantity $\mathcal{E}(\mathcal{S})$ is minimized over all n-subsets of $\mathcal{S}$ when the nodes in $\mathcal{S}$ are evenly distributed on the cycle $\mathcal{C}_N$.

Question: What about $t \neq 1$?

Example. Consider the cycle $\mathcal{C}_{10}$ and subsets $\mathcal{S}$ satisfying $|\mathcal{S}| = 5$. Consider the cases $t = 0, 1$ and 10.

Example. Consider the cycle $\mathcal{C}_{10}$ and subsets $\mathcal{S}$ satisfying $|\mathcal{S}| = 5$. Consider the cases $t = 0, 1$ and 10.

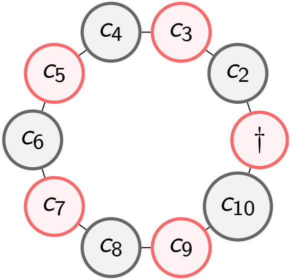


Figure : All maps are evenly distributed.

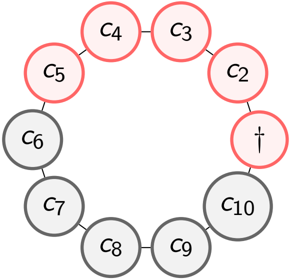


Figure : All maps are in one cluster.

	1	2	3	4	5	6	7	8	9	10	$\mathcal{E}(\mathcal{S})$
1	1	0	1	0	1	0	1	0	1	0	2.90
2	1	1	0	1	0	1	0	1	0	0	3.12
3	1	1	0	1	0	1	0	0	1	0	3.16
4	1	1	0	1	1	0	0	1	0	0	3.22
5	1	1	1	0	1	0	0	1	0	0	3.28
6	1	1	0	1	0	0	1	1	0	0	3.30
7	1	1	1	0	0	1	0	1	0	0	3.38
8	1	1	1	0	1	0	0	0	1	0	3.46
9	1	1	0	1	1	0	0	0	1	0	3.54
10	1	1	1	0	0	0	1	0	1	0	3.56
11	1	1	0	1	0	1	1	0	0	0	3.62
12	1	1	1	1	0	0	1	0	0	0	3.72
13	1	1	1	0	0	1	1	0	0	0	3.84
14	1	1	1	1	0	1	0	0	0	0	4.28
15	1	1	1	0	1	1	0	0	0	0	4.32
16	0	1	1	1	1	1	0	0	0	0	5.30

Table : Values of $\mathcal{E}(\mathcal{S})$ for $N = 10$, $n = 5$ and $t = 1$. One representative set is given for each of the sixteen distinct values of $\mathcal{E}(\mathcal{S})$.

	1	2	3	4	5	6	7	8	9	10	$\mathcal{E}(\mathcal{S})$
1	1	0	1	0	1	0	1	0	1	0	0.5
2	1	1	0	1	0	1	0	1	0	0	0.7
3	1	1	1	0	1	0	0	1	0	0	0.9
4	1	1	1	0	1	0	1	0	0	0	1.2
5	1	1	1	1	0	0	1	0	0	0	1.4
6	1	1	1	1	0	1	0	0	0	0	2.1
7	0	1	1	1	1	1	0	0	0	0	3.5

Table : Values of $\mathcal{E}(\mathcal{S})$ for $N = 10$, $n = 5$ and $t = 0$. One representative set is given for each of the seven distinct values of $\mathcal{E}(\mathcal{S})$.

	1	2	3	4	5	6	7	8	9	10	$\mathcal{E}(\mathcal{S})$
1	0	1	1	1	1	1	0	0	0	0	21.5
2	1	1	1	0	1	0	0	0	1	0	23.8
3	1	1	1	1	0	1	0	0	0	0	23.9
4	1	1	0	1	1	0	0	1	0	0	24.1
5	1	1	1	0	1	1	0	0	0	0	24.3
6	1	0	1	0	1	0	1	0	1	0	24.5
7	1	1	1	1	0	0	1	0	0	0	24.6
8	1	1	1	0	1	0	0	1	0	0	24.7
9	1	1	1	0	1	0	1	0	0	0	24.8
10	1	1	0	1	0	1	0	1	0	0	24.9
11	1	1	0	1	0	1	0	0	1	0	25.3
12	1	1	0	1	0	1	1	0	0	0	25.4
13	1	1	1	0	0	1	0	1	0	0	25.7
14	1	1	1	0	0	1	1	0	0	0	25.8

Table : Values of $\mathcal{E}(\mathcal{S})$ for $N = 10$, $n = 5$ and $t = 10$. One representative set is given for each of the sixteen distinct values of $\mathcal{E}(\mathcal{S})$.

Large delay factors, t, can greatly influence cluster detection. This will be evident for the two-dimensional grid.

- Consider a two-dimensional square toroidal grid on $m \times m$ nodes (i.e. $N = m^2$), $G(V, E) = [(i, j)]$ with (i, j) representing the vertex in the ith row and jth column of the grid.

- Consider a two-dimensional square toroidal grid on $m \times m$ nodes (i.e. $N = m^2$), $G(V, E) = [(i, j)]$ with (i, j) representing the vertex in the ith row and jth column of the grid.
- For convenience, we map the pair (i, j) to one dimension, via

$$\phi(i, j) = (i - 1)\, m + j, \qquad \textit{for } i, j = 1, 2, \ldots, m.$$

- Consider a two-dimensional square toroidal grid on $m \times m$ nodes (i.e. $N = m^2$), $G(V, E) = [(i, j)]$ with (i, j) representing the vertex in the ith row and jth column of the grid.
- For convenience, we map the pair (i, j) to one dimension, via

$$\phi(i, j) = (i - 1)\, m + j, \qquad \textit{for } i, j = 1, 2, \ldots, m.$$

- Similar to before, let $e_i = E[T | \mathcal{S}, \dagger, \phi^{-1}(i)]$ denote the expected amount of time for a walker starting at $\phi^{-1}(i)$ to reach the target $\dagger$. Suppose the set $\mathcal{S} = \{s_1, s_2, \ldots, s_n\}$ contains the designated maps while $\dagger \in \mathcal{S}$ is the target.

It is worth noting that if the pair (i, j) satisfies either i or j is equal to 1 or m, then by the toroidal effect, it still has four neighbors.

It is worth noting that if the pair (i, j) satisfies either i or j is equal to 1 or m, then by the toroidal effect, it still has four neighbors.

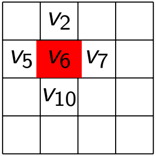


Figure : Neighbors of v_6.

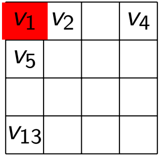


Figure : Neighbors of v_1.

For convenience in what follows, we will at times refer to the pair $\phi^{-1}(i)$ simply as i or potentially v_i.

Define the ratio of the expected amount of time to hit the node s_i, as target, with and without the removal of s_j from $\mathcal{S}$ via,

Define the ratio of the expected amount of time to hit the node s_i, as target, with and without the removal of s_j from $\mathcal{S}$ via,

$$r_{i,j} = \frac{\left(\bar{e}\left(\left(\mathcal{S} \setminus s_j\right), s_i\right)\right)}{\left(\bar{e}\left(\mathcal{S}, s_i\right)\right)} \qquad i, j = 1, 2, \ldots, n.$$

Define the ratio of the expected amount of time to hit the node s_i, as target, with and without the removal of s_j from $\mathcal{S}$ via,

$$r_{i,j} = \frac{\left(\bar{e}\left(\left(\mathcal{S} \setminus s_j\right), s_i\right)\right)}{\left(\bar{e}\left(\mathcal{S}, s_i\right)\right)} \qquad i, j = 1, 2, \ldots, n.$$

We refer to the matrix $\mathcal{V} = \mathcal{V}_{\mathcal{S}} = [r_{i,j}]$ as the *(transport) value matrix*.

Define the ratio of the expected amount of time to hit the node s_i, as target, with and without the removal of s_j from $\mathcal{S}$ via,

$$r_{i,j} = \frac{\left(\bar{e}\left(\left(\mathcal{S} \setminus s_j\right), s_i\right)\right)}{\left(\bar{e}\left(\mathcal{S}, s_i\right)\right)} \qquad i,j = 1,2,\ldots,n.$$

We refer to the matrix $\mathcal{V} = \mathcal{V}_{\mathcal{S}} = [r_{i,j}]$ as the *(transport) value matrix*.

Consider the ratios (of ratios),

$$\rho_{i,j} \overset{def}{=} \frac{r_{i,j}}{r_{j,i}}.$$

Define the ratio of the expected amount of time to hit the node s_i, as target, with and without the removal of s_j from $\mathcal{S}$ via,

$$r_{i,j} = \frac{\left(\bar{e}\left(\left(\mathcal{S} \setminus s_j\right), s_i\right)\right)}{\left(\bar{e}\left(\mathcal{S}, s_i\right)\right)} \qquad i, j = 1, 2, \ldots, n.$$

We refer to the matrix $\mathcal{V} = \mathcal{V}_{\mathcal{S}} = [r_{i,j}]$ as the *(transport) value matrix*.

Consider the ratios (of ratios),

$$\rho_{i,j} \stackrel{def}{=} \frac{r_{i,j}}{r_{j,i}}.$$

We refer to the matrix $[\rho_{ij}]$ as the *value ratio matrix*, $\mathcal{V}_{rat}$ and denote this by

$$\mathcal{V}_{rat} \stackrel{def}{=} [\rho_{i,j}] = \left[\frac{r_{i,j}}{r_{j,i}}\right].$$

Summarize the information contained in the matrix $\mathcal{V}_{rat}$ by summing over columns to obtain "importance" values, w_j, i.e.

Summarize the information contained in the matrix $\mathcal{V}_{rat}$ by summing over columns to obtain "importance" values, w_j, i.e.

$$w_j \stackrel{def}{=} \frac{\sum_{i=1}^{n} \rho_{i,j}}{n}.$$

Summarize the information contained in the matrix $\mathcal{V}_{rat}$ by summing over columns to obtain "importance" values, w_j, i.e.

$$w_j \stackrel{def}{=} \frac{\sum_{i=1}^{n} \rho_{i,j}}{n}.$$

We refer to w_j as the jth *shading value*.

Summarize the information contained in the matrix $\mathcal{V}_{rat}$ by summing over columns to obtain "importance" values, w_j, i.e.

$$w_j \stackrel{def}{=} \frac{\sum_{i=1}^{n} \rho_{i,j}}{n}.$$

We refer to w_j as the jth *shading value*.
The greater w_j is, the more valuable s_j is to the other maps in $\mathcal{S}$, as a whole.

Two-Dimensional Toroidal Grid Examples

Suppose we have a two-dimensional square toroidal grid with $N = 100$ (10×10), and that $\mathcal{S} \in V$ is the set of maps. Consider cases where $t = 1, 100$. Plots with maps shaded via the ranks of the shading values are given below; all the shaded locations represent maps. The darker the color is, the smaller the corresponding shading value is. The overall quantity $\mathcal{E}(\mathcal{S})$ is also included.

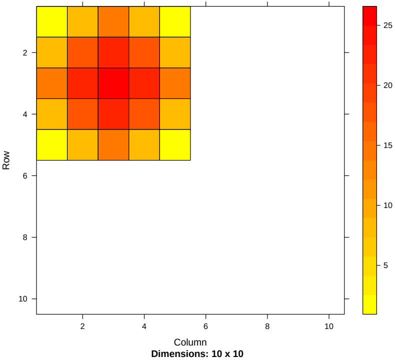


Figure : Here, $t = 1$, $\mathcal{E}(\mathcal{S}) = 21.13$.

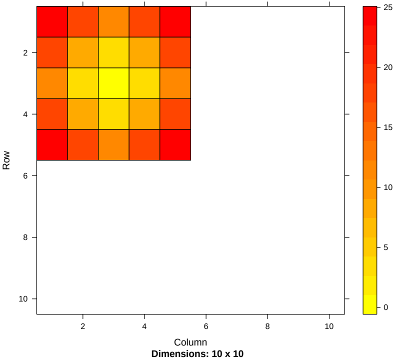


Figure : Here, $t = 100$, $\mathcal{E}(\mathcal{S}) = 369.86$.

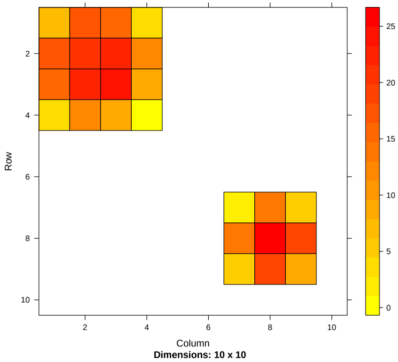


Figure : Here, $t = 1$, $\mathcal{E}(\mathcal{S}) = 15.69$.

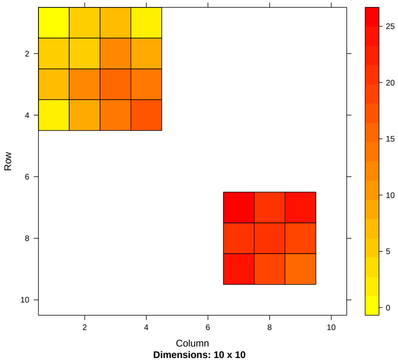


Figure : Here, $t = 100$, $\mathcal{E}(\mathcal{S}) = 498.27$.

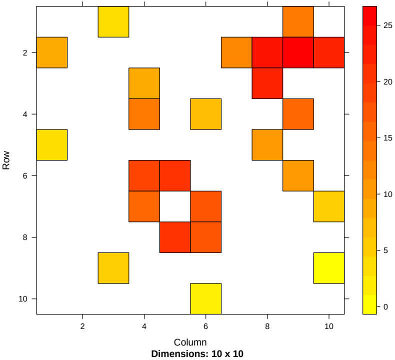


Figure : Here, $t = 1$, $\mathcal{E}(\mathcal{S}) = 8.62$.

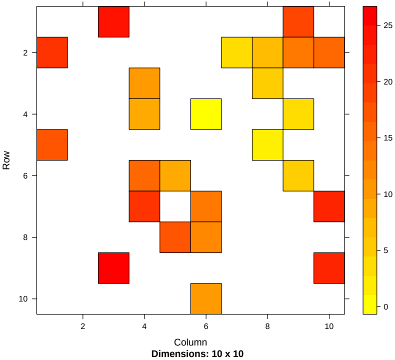


Figure : Here, $t = 100$, $\mathcal{E}(\mathcal{S}) = 501.07$.

Some additional plots for the case $t = 1$ with $n = 34$ maps. Note the change in $\mathcal{E}(\mathcal{S})$ as the maps become less "clustered".

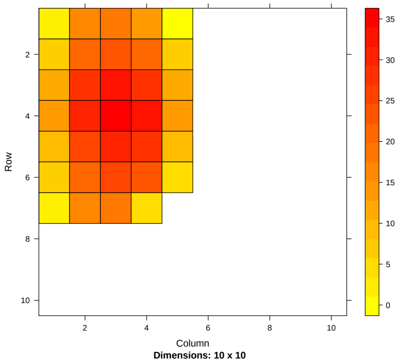


Figure : Here, $t = 1$, $\mathcal{E}(\mathcal{S}) = 15.04$.

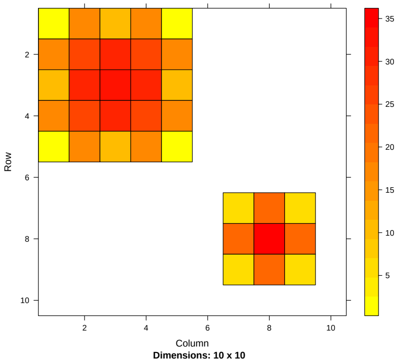


Figure : Here, $t = 1$, $\mathcal{E}(\mathcal{S}) = 11.66$.

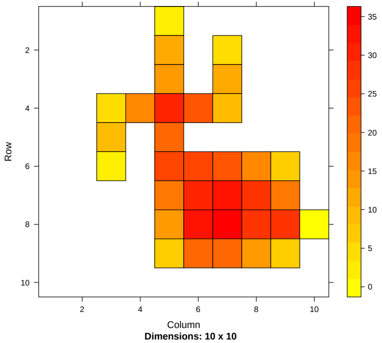


Figure : Here, $t = 1$, $\mathcal{E}(\mathcal{S}) = 11.61$.

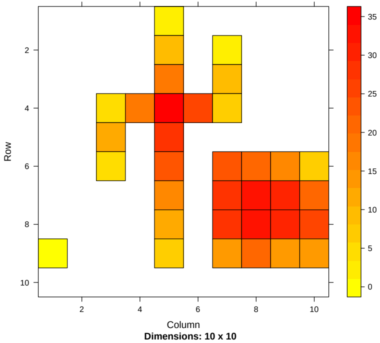


Figure : Here, $t = 1$, $\mathcal{E}(\mathcal{S}) = 10.44$.

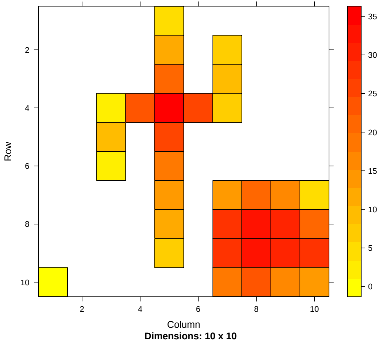


Figure : Here, $t = 1$, $\mathcal{E}(\mathcal{S}) = 10.24$.

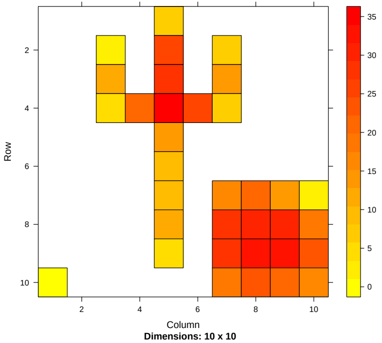


Figure : Here, $t = 1$, $\mathcal{E}(\mathcal{S}) = 10.04$.

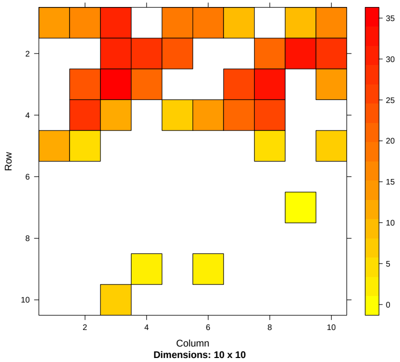


Figure : Here, $t = 1$, $\mathcal{E}(\mathcal{S}) = 9.42$.

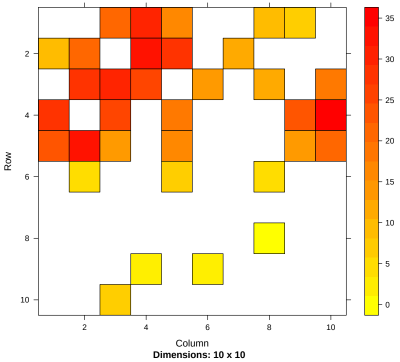


Figure : Here, $t = 1$, $\mathcal{E}(\mathcal{S}) = 8.44$.

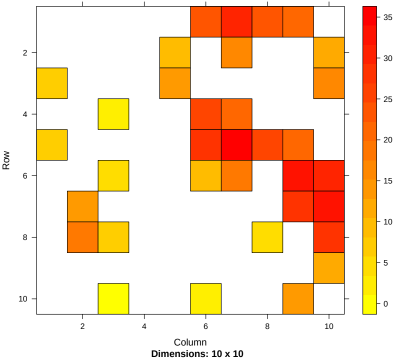


Figure : Here, $t = 1$, $\mathcal{E}(\mathcal{S}) = 7.76$.

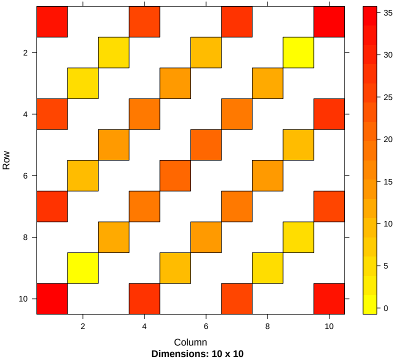


Figure : Here, $t = 1$, $\mathcal{E}(\mathcal{S}) = 6.48$.

Returning to the Cycle Case.

Recall $\{X_0, X_1, \ldots, X_{T_0}, \ldots\}$ be a random walk starting at $X_0 = x_0$. We then have

$$\begin{aligned} T_0 &= \min\{i \geq 0 | X_i \in \mathcal{S}\}, \\ T_1 &= dist(X_{T_0}, \dagger) \qquad \text{and} \\ T &= T_0 + tT_1. \end{aligned}$$

where $t \geq 0$ is a delay factor.

Recall for $s_1, s_2 \in \mathcal{S}$, where $s_1 < s_2$, set

$$\mathcal{I} = \mathcal{I}_{s_1,s_2} \overset{def}{=} (s_1, s_2) = \{s_1 + 1, s_1 + 2, \ldots, s_2 - 1\},$$

$$a \overset{def}{=} dist(s_1, \dagger) \quad \text{and} \quad b \overset{def}{=} dist(s_2, \dagger),$$

and

$$K_t \overset{def}{=} \frac{1}{|\mathcal{I}|} \sum_{v_i \in \mathcal{I}} e\left(\mathcal{S}, \dagger, v_i\right).$$

Minimization with $t \neq 1$

Lemma

Suppose $\mathcal{I} \cap \mathcal{S} = \emptyset$ *and* $t \geq 0$*, is the delay factor. Then, with* $d = |\mathcal{I}| + 1$*,*

$$K_t = \frac{d^2 + d}{6} + \frac{(a+b)t}{2}.$$

Minimization with $t \neq 1$

Lemma

Suppose $\mathcal{I} \cap \mathcal{S} = \emptyset$ and $t \geq 0$, is the delay factor. Then, with $d = |\mathcal{I}| + 1$,

$$K_t = \frac{d^2 + d}{6} + \frac{(a+b)t}{2}.$$

Lemma

Suppose $\mathcal{I} \cap \mathcal{S} = s^*$, $r = dist(s^*, s_1)$ and $t \geq 0$, is the delay factor. Then K_t is minimized when

$$r = \frac{d}{2} + \frac{(b-a-d)t}{2d}.$$

Maximization

Maximization

Maximization for K_t

Maximization

Maximization for K_t

Lemma

Suppose $\mathcal{I} \cap \mathcal{S} = s^*$, $r = dist(s^*, s_1)$ and $t \geq 0$, is the delay factor. Then K_t is maximized when

$$r = \begin{cases} d-1 & \text{if } t \geq \frac{2}{d} \\ 1 & \text{if } t \leq \frac{2}{d} \end{cases}$$

Maximization

Maximization for K_t

Lemma

Suppose $\mathcal{I} \cap \mathcal{S} = s^*$, $r = dist(s^*, s_1)$ and $t \geq 0$, is the delay factor. Then K_t is maximized when

$$r = \begin{cases} d-1 & \text{if } t \geq \frac{2}{d} \\ 1 & \text{if } t \leq \frac{2}{d} \end{cases}$$

Corollary

Suppose $\mathcal{I} \cap \mathcal{S} = s^*$, $r = dist(s^*, s_1)$. Then K_1 is maximized when $r = d-1$.

Example

$\dagger$	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}	c_{13}	$\bar{e}(\mathcal{S}, c_1)$
1		1						1	1			1	6.03

Example

†	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}	c_{13}	$\bar{e}(\mathcal{S}, c_1)$
1		1						1	1			1	6.03

†	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}	c_{13}	$\bar{e}(\mathcal{S}, c_1)$
1							1	1	1			1	7.34

Example

$\dagger$	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}	c_{13}	$\bar{e}(\mathcal{S}, c_1)$
1		1						1	1			1	6.03

$\dagger$	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}	c_{13}	$\bar{e}(\mathcal{S}, c_1)$
1							1	1	1			1	7.34

$\dagger$	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}	c_{13}	$\bar{e}(\mathcal{S}, c_1)$
1							1	1	1	1			7.34

Lemma

Suppose on a cycle $\mathcal{C}_N$, $\mathcal{S}_1$ and $\mathcal{S}_2$ are two differing map sets. In addition, suppose $P \notin \mathcal{I}$,

$$1 \leq l < k < s_2 - s_1 - 1 = b - a - 1 = d - 1,$$

$$\mathcal{I} \cap \mathcal{S}_1 = \{s_1 + 1, s_1 + 2, \dots, s_1 + k\}$$

and

$$\mathcal{S}_2 \cap \mathcal{I} = \{s_1 + 1, \dots, s_1 + l\} \cup \{s_2 - (k - l), \dots, s_2 - 1\}.$$

Then K_t thought of as a function of map sets satisfies,

$$K_t(\mathcal{S}_1) = K_t(\mathcal{S}_2).$$

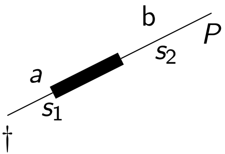


Figure : The height indicates the distance to †. The bold parts indicate maps $(\mathcal{S}_1 \cap \mathcal{I})$.

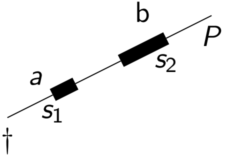


Figure : The height indicates the distance to †. The bold parts indicate maps $(\mathcal{S}_2 \cap \mathcal{I})$.

Lemma

Suppose $\mathcal{I} \cap \mathcal{S} = \{s^*, s^* + 1, ..., s^* + k - 1\}$ *and* $dist(s_1, s^*) = r$. *Then* $K_1(r)$ *is maximized when* $r = d - k$, *where* $d = |\mathcal{I}| + 1$.

Proof.

Proof. Suppose N is even (the case N is odd is similar). Assume $b \geq a$ and denote by P, the furthest vertex from $\dagger$. Set $p = dist(P, \dagger)$.

Proof. Suppose N is even (the case N is odd is similar). Assume $b \geq a$ and denote by P, the furthest vertex from $\dagger$. Set $p = dist(P, \dagger)$.

Case 1: Assume $b = a + d$.

Proof. Suppose N is even (the case N is odd is similar). Assume $b \geq a$ and denote by P, the furthest vertex from $\dagger$. Set $p = dist(P, \dagger)$.

Case 1: Assume $b = a + d$.

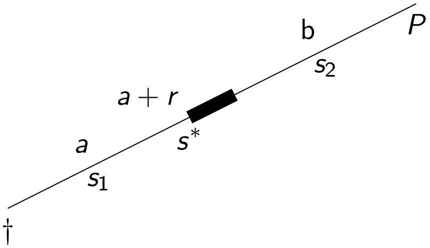


Figure : The case $b = a + d$. The height is intended to indicate the distance to $\dagger$. The bold lines indicate map clusters.

Case 2: Assume $b < a + d$.

Case 2: Assume $b < a + d$.

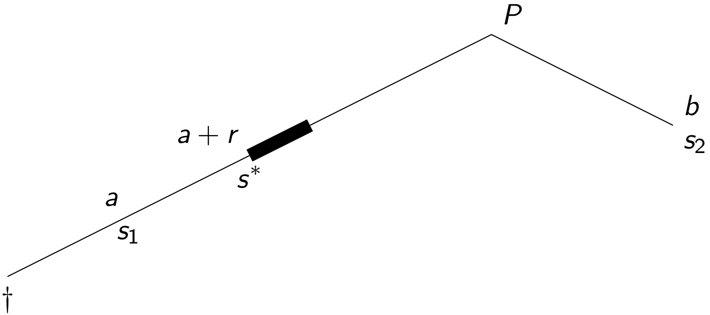


Figure : The case $b \leq d + a - 1$. The height is intended to indicate distance to $\dagger$. Here, the whole cluster lies on the left arm. The bold line indicates maps.

Subcase 1: Neither arm is able to hold the whole cluster.

Subcase 1: Neither arm is able to hold the whole cluster.

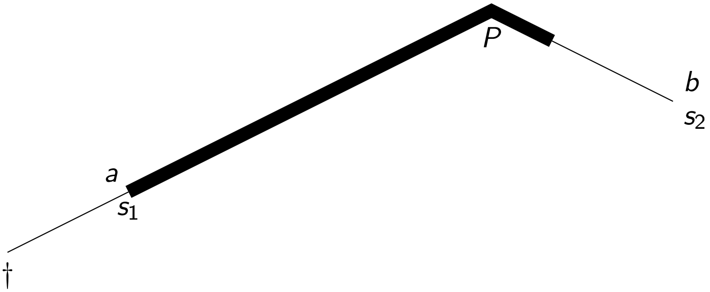


Figure : The case $b \leq d + a - 1$. The height is intended to indicate distance to $\dagger$. The bold lines indicate map clusters.

Subcase 2: The short arm is not able to hold the whole cluster.

Subcase 2: The short arm is not able to hold the whole cluster.

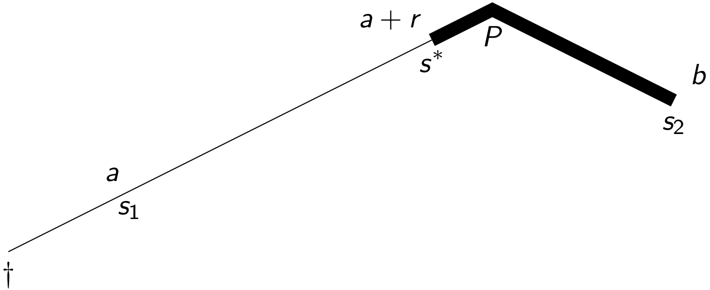


Figure : The case $b \leq d + a - 1$. The height is intended to indicate distance to $\dagger$. The bold lines indicate map clusters.

Subcase 3: Both arms are long enough to hold the whole cluster.

Subcase 3: Both arms are long enough to hold the whole cluster.

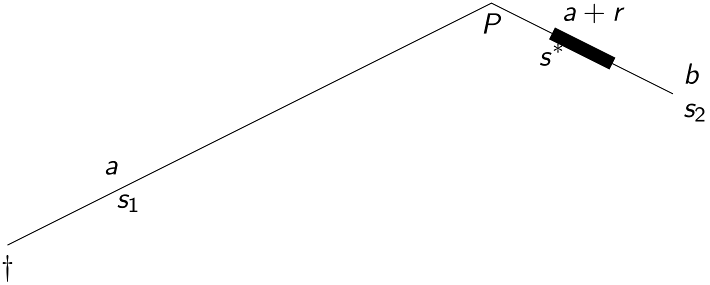


Figure : The case $b \leq d + a - 1$. The height is intended to indicate distance to $\dagger$. The bold lines indicate a map cluster.

Theorem

Suppose $0 \leq t \leq 1$. The quantity $\mathcal{E}(\mathcal{S})$ is maximized over all n-subsets of $\mathcal{S}$ when the nodes in $\mathcal{S}$ are consecutively clustered on the cycle $\mathcal{C}_N$.

Proof

- Suppose $\mathcal{S} = \{s_1, \ldots, s_n\}$ is a subset of $\mathcal{C}_N$, which does not consist of consecutive nodes in $\mathcal{C}_N$.

Proof

- Suppose $\mathcal{S} = \{s_1, \ldots, s_n\}$ is a subset of $\mathcal{C}_N$, which does not consist of consecutive nodes in $\mathcal{C}_N$.
- For each $1 \leq i \leq n$, we employ the preceding lemmas to construct a chain of n-sets, $\mathcal{S}_{i,1}, \mathcal{S}_{i,2}, \ldots, \mathcal{S}_{i,\nu_i}$, with the last, S_{i,ν_i}, consisting of two disjoint non-neighboring sets of consecutive nodes in $\mathcal{C}_N$, satisfying

  $$\bar{e}(\mathcal{S}, s_i) \leq \bar{e}(\mathcal{S}_{i,1}, s_i) \leq \cdots \leq \bar{e}(\mathcal{S}_{i,\nu_i}, s_i) \leq \bar{e}(\{1, 2, \ldots, n\}, i)$$

  where $\mathcal{S}_\nu$ is comprised of consecutive nodes in $\mathcal{C}_N$.

- Finally,

$$\begin{aligned}
\mathcal{E}(\mathcal{S}) &= \frac{1}{n}\sum_{1\leq i\leq n}\bar{e}(\mathcal{S}, s_i) \\
&\leq \frac{1}{n}\sum_{1\leq i\leq n}\bar{e}(\{1,2,\ldots,n\}, i) \\
&\leq \mathcal{E}(\{1,2,\ldots,n\}),
\end{aligned}$$

and the theorem is proven.

Example

$\dagger$	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}	c_{13}	$\bar{e}(\mathcal{S}, c_1)$
1		1						1	1			1	6.03

Example

$\dagger$	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}	c_{13}	$\bar{e}(\mathcal{S}, c_1)$
1		1						1	1			1	6.03

$\dagger$	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}	c_{13}	$\bar{e}(\mathcal{S}, c_1)$
1							1	1	1			1	7.34

Example

†	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}	c_{13}	$\bar{e}(\mathcal{S}, c_1)$
1		1						1	1			1	6.03

†	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}	c_{13}	$\bar{e}(\mathcal{S}, c_1)$
1							1	1	1			1	7.34

†	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}	c_{13}	$\bar{e}(\mathcal{S}, c_1)$
1							1	1	1	1			7.34

Example

$\dagger$	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}	c_{13}	$\bar{e}(\mathcal{S}, c_1)$
1		1						1	1			1	6.03

$\dagger$	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}	c_{13}	$\bar{e}(\mathcal{S}, c_1)$
1							1	1	1			1	7.34

$\dagger$	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}	c_{13}	$\bar{e}(\mathcal{S}, c_1)$
1							1	1	1	1			7.34

$\dagger$	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}	c_{13}	$\bar{e}(\mathcal{S}, c_1)$
1	1	1	1	1									11.08

Thank you!