ESCAPE OF RESOURCES IN A DISTRIBUTED CLUSTERING PROCESS

JACOB VAN DEN BERG
CWI and VU University Amsterdam
email: J.van.den.Berg@cwi.nl

MARCELO R. HILÁRIO[1]
IMPA, Rio de Janeiro
email: marcelo@impa.br

ALEXANDER E. HOLROYD[2]
Microsoft Research and University of British Columbia
email: holroyd@math.ubc.ca

Abstract

In a distributed clustering algorithm introduced by Coffman, Courtois, Gilbert and Piret [1], each vertex of $\mathbb{Z}^d$ receives an initial amount of a resource, and, at each iteration, transfers all of its resource to the neighboring vertex which currently holds the maximum amount of resource. In [4] it was shown that, if the distribution of the initial quantities of resource is invariant under lattice translations, then the flow of resource at each vertex eventually stops almost surely, thus solving a problem posed in [2]. In this article we prove the existence of translation-invariant initial distributions for which resources nevertheless escape to infinity, in the sense that the the final amount of resource at a given vertex is strictly smaller in expectation than the initial amount. This answers a question posed in [4].

1 Introduction

1.1 Definitions and statement of the main result

Consider, for $d \geq 1$, the d-dimensional integer lattice. This is the graph with vertex set $\mathbb{Z}^d$, and edge set comprising all pairs of vertices (x, y) $(= (y, x))$ with $|x - y| = 1$. Here $|\cdot|$ denotes the 1-norm. We use the notation $\mathbb{Z}^d$ for this graph as well as for its vertex set. It will be clear from the context which of the two is meant.

The following model for '*distributed clustering*' was introduced by Coffman, Courtois, Gilbert and Piret [1]. To each vertex x of the lattice $\mathbb{Z}^d$, we assign a random nonnegative number $C_0(x) \in [0, \infty]$ which we regard as the initial amount of a 'resource' placed at x at time 0. (The family $(C_0(x); x \in \mathbb{Z}^d)$ is not necessarily assumed independent). Then we define a discrete-time evolution in which, at each step, each vertex transfers its resource to the 'richest' neighbouring vertex. More precisely, the evolution is defined recursively as follows. Suppose that, at time n, the *amount of resource* at each vertex x is $C_n(x)$. Let $N(x) = \{y \in \mathbb{Z}^d : |x - y| \leq 1\}$ be the *neighbourhood* of x (note that it includes x itself) and define

$$M_n(x) = \left\{ y \in N(x) : \ C_n(y) = \max_{z \in N(x)} C_n(z) \right\}.$$

Now let $\nu_n(x)$ be a vertex chosen uniformly at random in $M_n(x)$, independently for each x, and take:

$$a_n(x) = \left\{ \begin{array}{ll} x, & \text{if } C_n(x) = 0 \\ \nu_n(x), & \text{if } C_n(x) > 0. \end{array} \right.$$

Finally, define

$$C_{n+1}(x) := \sum_{y \, : \, a_n(y) = x} C_n(y).$$

For a fixed vertex x, the random variable $C_0(x)$ will be called the *initial amount of resource* at x, and the family $(C_0(x); x \in \mathbb{Z}^d)$ will be called the *initial configuration*. Analogously, $(C_n(x); x \in \mathbb{Z}^d)$ will be called the *configuration at time n*. Note that $a_n(x)$ is the vertex to which the resources located at x at time n (if any) will be transferred during the $(n+1)$-th step of the evolution. We say that there is a *tie* in x at time n if $C_n(x) > 0$ and the cardinality of $M_n(x)$ is strictly greater than one. In case this occurs, $a_n(x)$ is chosen uniformly at random among the vertices around x that maximize C_n. Note that, apart from those possible tie breaks, all the randomness is contained in the initial configuration. As soon as a vertex has zero resource, its resource remains zero forever. Also note that, when two or more vertices transfer their resources to the same vertex, these resources are added up. Thus this algorithm models a clustering process in the lattice starting from a disordered initial configuration.

For a fixed vertex x, we use the notation $C_\infty(x)$ for $\lim_{n \to \infty} C_n(x)$ in case this limit exists. We write $\mathbb{E}$ for expectation with respect to the underlying probability measure.

Our main result is the following theorem. The proof is given in Section 3.

Theorem 1. *Let $d \geq 2$. There exists a translation-invariant distribution for the initial configuration $(C_0(x); x \in \mathbb{Z}^d)$ such that, for each $x \in \mathbb{Z}^d$,*

$$\mathbb{E}\left[C_\infty(x)\right] < \mathbb{E}\left[C_0(x)\right]. \tag{1}$$

1.2 Background and motivation

Here is some more terminology. If, for all sufficiently large n, we have that $a_n(x) = x$ and $a_n(y) \neq x$ for all neighbours y of x, then we say that the flow at x *terminates* after finitely many steps. In that case, the limit $C_\infty(x)$ is attained after finitely many iterations and will be called the *final amount of resource at x*. If for all sufficiently large n we have $a_{n+1}(x) = a_n(x)$, then we say that x eventually transfers its resource to the same fixed vertex.

The following stability questions for this process (formulated here similarly as in [4]) have been investigated in the literature:

Question 1. *Does each vertex eventually transfer its resource to the same fixed vertex almost surely?*

Question 2. *Does the flow at each vertex terminate after finitely many steps almost surely?*

Question 3. *If the answer to the previous question is affirmative, is the expected final amount of resource of a vertex equal to the expected initial amount?*

Of course the answers to the above questions may depend on the assumptions made about the distribution of the initial configuration. Note that if the answer to Question 2 is affirmative, then so is the answer to Question 1. In that case, answering Question 3 is equivalent to answering the question whether the resource quantity that started on a given vertex will eventually stop moving almost surely. So, informally, Question 2 is related to fixation while Question 3 is related to conservation.

Van den Berg and Meester [2] considered the case $d = 2$ and i.i.d. initial resource quantities. Using translation-invariance and symmetries of the system they proved that the answer to Question 1 is positive in the case that the initial quantities of resource have a continuous distribution. They also showed that, if the resources are integer valued, then Question 2 has a positive answer as well.

Later, van den Berg and Ermakov [3] considered again i.i.d. continuously distributed initial quantities of resource on the two-dimensional lattice. Using a percolation approach, they were able to relate Questions 2 and 3 to a finite (but large) computation. By using Monte Carlo simulation, they obtained overwhelming evidence that the answer to these questions is positive for this case.

In [4] it was proved that, for every dimension and every translation-invariant distribution of the initial configuration, the answer to Question 2 is positive. However, Question 3 was left open. Our Theorem 1 says that, for some initial distributions in this class, the answer to that question is negative.

The conclusion of Theorem 1 is false for $d = 1$. To see that, suppose that the probability that the resource starting at the origin does not stop after finitely many steps is positive. Then, by translation invariance, there is, with positive probability, a positive density of vertices for which the initial resource will not stop after finite time. This implies that, with positive probability, there are infinitely many steps at which resource enters or leaves the origin, contradicting the fixation result of [4] mentioned in the previous paragraph. This argument can be generalized for example to any graph of the form $\mathbb{Z} \times G$, where G is a finite vertex-transitive graph. (For such graphs translation-invariance is replaced with automorphism-invariance).

In order to prove Theorem 1, we will construct a random collection (forest) of one-ended trees, which is embedded in $\mathbb{Z}^d$, in a translation-invariant way, and then assign resource quantities to the vertices in such a way that, during the evolution, each resource follows the unique infinite self-avoiding path to infinity in the forest. In Section 2 we present a short discussion about the existence of certain random forests on $\mathbb{Z}^d$. In Section 3, Theorem 1 is proved. In Section 4 we present some concluding remarks and open questions.

2   Translation-invariant forests on $\mathbb{Z}^d$

Let G be an infinite graph. A *forest* of G is a subgraph of G that has no cycles. A *tree* is a connected forest. A subgraph *spans* G if it contains every vertex of G. A *spanning forest* (respectively *tree*) on G is a subgraph of G that is a forest (respectively a tree) and that spans G. The *leaves* of a forest T are the vertices of T that only have one neighbor in the forest. The *number of ends* of a tree is the number of distinct self-avoiding infinite paths starting from a given vertex. A tree is said to be *one-ended* if it has one end.

We choose the d-dimensional integer lattice as the underlying graph. For this choice, the literature provides several constructions of random spanning forests with translation-invariant distributions, for example, the uniform spanning tree [5], and the minimal spanning tree [6]. To be explicit, we briefly discuss one construction, based on the two-dimensional minimal spanning tree.

Let E be the set of edges of the lattice $\mathbb{Z}^2$, and let $\{U_e; e \in E\}$ be a family of independent random variables distributed uniformly in the interval $[0, 1]$. For each cycle of the lattice, delete the edge having the maximum U-value on the cycle. The resulting random graph is called *(free) minimal spanning forest* and is known to be almost surely a one-ended tree which is invariant and ergodic under lattice translations (see [7]).

For $d > 2$, we can use the two-dimensional minimal spanning forest to construct a random forest in $\mathbb{Z}^d$ of which the distribution is invariant under lattice translations, and of which every component is one-ended. We regard $\mathbb{Z}^d$ as $\mathbb{Z}^2 \times \mathbb{Z}^{d-2}$ and in each 'layer' $\mathbb{Z}^2 \times \{z\}$ (where z runs over $\mathbb{Z}^{d-2}$) we embed an independent copy of the two-dimensional minimal spanning tree T_z. The resulting subgraph of $\mathbb{Z}^d$ is a translation-invariant random spanning forest with one-ended components. This gives the following lemma.

Lemma 2. *For each $d \geq 2$ there exists a translation-invariant random spanning forest on $\mathbb{Z}^d$, of which each connected component is one-ended almost surely.*

Corollary 3. *For each $d \geq 2$ there exists a translation-invariant random forest T on $\mathbb{Z}^d$, for which the following two properties hold almost surely.*
(i) Every connected component of T is one-ended.
(ii) Every edge of $\mathbb{Z}^2$ of which both endpoints are in T is an edge of T.

Proof. Let H be a spanning forest as in Lemma 2 and write F for the set of its edges, and V for the set of its vertices. Let $\widetilde{H}$ be the forest with vertex set $\{2x : x \in V\} \cup \{x + y : (x, y) \in F\}$ and edge set $E = \{(2x, x + y) : (x, y) \in F\}$. Informally, $\widetilde{H}$ corresponds to the forest which is obtained when H is scaled up by a factor 2. Thus to each edge (x, y) of H, there correspond two edges, $(2x, x + y)$ and $(x + y, 2y)$, in $\widetilde{H}$. Note that $\widetilde{H}$ is a random forest which is invariant under translations of $2\mathbb{Z}^d$, and which has the property that every pair x, y of its vertices satisfying $|x - y| = 1$ is connected by an edge of $\widetilde{H}$. To restore invariance under *all* translations of $\mathbb{Z}^d$, let W be a uniformly random element of the discrete cube $\{0, 1\}^d$, independent of $\widetilde{H}$, and set $T = \widetilde{H} + W$. $\square$

3 Proof of main result

In this section we fix $d \geq 2$. We will prove Theorem 1 by giving an explicit construction of an initial configuration $(C_0(x), x \in \mathbb{Z}^d)$ whose distribution is translation-invariant and for which (1) holds.

Let T be a random forest on $\mathbb{Z}^d$ as given by Corollary 3. For vertices x and y of T we write $x \sim y$ if (x, y) is an edge of T. We define a (random) partial order $\leq$ on $\mathbb{Z}^d$ by setting $y \leq x$ if and only if x and y are vertices of T and x belongs to the unique infinite self-avoiding path in T starting at y. If $y \leq x$ we say that x is an *ancestor* of y and that y is a *descendant* of x. If $y \leq x$ and $x \sim y$ we say that x is a *parent* of y and that y is a *child* of x. Note that every vertex of T has a unique parent. Moreover, for every vertex x of T, exactly one vertex in $\{y : y \sim x\}$ is the parent of x, and the others are descendants of x.

We now define, for each $x \in \mathbb{Z}^d$, the initial quantity of resource at x by:

$$C_0(x) = \begin{cases} \sum_{y \in \mathbb{Z}^d} \mathbf{1}[y \leq x], & \text{if } x \in T \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Note that, if $x \in T$, then $C_0(x)$ is the number of descendants of x. Since every connected component of T is one-ended, it follows from the definitions that this number is finite. Also note that, since the distribution of T is invariant under the translations of $\mathbb{Z}^d$, so is that of the family $\left(C_0(x); x \in \mathbb{Z}^d\right)$.

We now define a nested ('decreasing') sequence of forests that will be shown to describe the dynamics of resources when $C_0(x)$ is given by (2). For a forest S, let $\phi(S)$ denote the forest obtained from S by deleting all its leaves. Let $T_0 = T$ and, for $n = 1, 2, \ldots$, define inductively $T_n = \phi(T_{n-1})$.

The following observation follows easily from the definitions.

Observation 4. *Let x be a vertex of T and $n \geq 0$. Then x is in T_{n+1} if and only if some child of x is in T_n.*

Lemma 5. *For every vertex x in T, there is a finite index n_0 (depending on x) such that, for all $n \geq n_0$, x does not belong to T_n.*

Proof. By Observation 4, $n_0(x)$ is at most 1 plus the number of descendents of x. As we mentioned before, this number is finite. $\qquad\square$

Lemma 6. *Suppose that, for all x, $C_0(x)$ is given by (2). Then for all $n \geq 0$,*

$$C_n(x) \begin{cases} > \sum_{y:y\sim x, y\leq x} C_n(y), & \text{if } x \in T_n; \\ = 0, & \text{if } x \notin T_n. \end{cases} \tag{3}$$

Proof. We use induction on n. To verify (3) for $n = 0$ we note that, if x belongs to $T_0(= T)$ then, by (2),

$$\sum_{\substack{y:y\sim x, \\ y\leq x}} C_0(y) = \sum_{\substack{y:y\sim x, \\ y\leq x}} \sum_{z\in\mathbb{Z}^d} \mathbf{1}[z \leq y] = \sum_{z\in\mathbb{Z}^d\setminus\{x\}} \mathbf{1}[z \leq x] = C_0(x) - 1.$$

Now, suppose that (3) holds for a given n. Since T was taken as in Corollary 3, two vertices of T_n which are adjacent in $\mathbb{Z}^d$ must be linked by an edge of T_n. By this and (3) it follows that, for each vertex z of T_n, $a_n(z)$ is the parent of z. Therefore, and because $C_n \equiv 0$ outside T_n, we have

$$C_{n+1}(x) = 0 \qquad\qquad\qquad\qquad\qquad\qquad \text{if } x \notin T_{n+1}; \tag{4}$$

$$C_{n+1}(x) = \sum_{y:y\sim x, y\leq x} C_n(y) \qquad\qquad\qquad \text{if } x \in T_{n+1}. \tag{5}$$

By applying (5), (3) and Observation 4 (and noting that (5) also holds for $x \notin T_{n+1}$, since then both sides of (5) are equal to 0), we get, for $x \in T_{n+1}$,

$$C_{n+1}(x) = \sum_{\substack{y:y\sim x, \\ y\leq x}} C_n(y) > \sum_{\substack{y:y\sim x, \\ y\leq x}} \sum_{\substack{z:z\sim y, \\ z\leq y}} C_n(z) = \sum_{\substack{y:y\sim x, \\ y\leq x}} C_{n+1}(y). \tag{6}$$

Now (4) and (6) complete the induction step, and the proof of Lemma 6. $\qquad\square$

Proof of Theorem 1. Let the initial configuration be defined as in (2). Let $x \in \mathbb{Z}^d$. By Lemma 6 and Lemma 5, we have that almost surely $C_n(x) = 0$ for all sufficiently large n. Hence $C_\infty(x) = 0$ almost surely. On the other hand, it is clear that $C_0(x) > 0$ with positive probability, and hence $\mathbb{E}[C_0(x)] > 0$. $\qquad\square$

4   Concluding remarks and open problems

At the end of the proof of Theorem 1 we mentioned the obvious fact that $\mathbb{E}[C_0(x)] > 0$ for every x. It turns out that this expectation is even ∞. Indeed, we have

$$\mathbb{E}[C_0(x)] = \sum_{y \in \mathbb{Z}^2} \mathbb{P}[y \leq x] = \sum_{y \in \mathbb{Z}^2} \mathbb{P}[x + y \leq x]$$
$$= \sum_{y \in \mathbb{Z}^2} \mathbb{P}[x \leq x - y] = \sum_{y \in \mathbb{Z}^2} \mathbb{P}[x \leq y] = \infty,$$

where the second and forth equality follow by relabeling, the third equality follows by translation-invariance and the last inequality follows from the fact that x has infinitely many ancestors almost surely.

We have not been able to construct an example where resources escape to infinity but the initial amount of resource at a given vertex has finite expectation. It is an interesting question whether such examples exist.

In particular, in our construction, the initial configuration was chosen in such a way that, almost surely, the induced dynamics takes place in a forest with one-ended components, embedded in $\mathbb{Z}^d$ and, at each step, the resources are transferred from every vertex with non-zero resource to its parent. It is not clear if for every initial configuration with these properties the expectation of the initial amount of resource of a vertex is infinite. We state these considerations more formally by the following two questions.

Question 4. *Suppose that $(C_0(x); x \in \mathbb{Z}^d)$ has a translation-invariant distribution and is positive exactly on the vertices of a forest with one-ended components. Furthermore, suppose that during the n-th step of the evolution, every vertex for which $C_{n-1}(x) > 0$ transfers its resource to its parent. Is it the case that $\mathbb{E}[C_0(x)] = \infty$?*

Question 5. *Does there exist a translation-invariant distribution for the initial configuration for which $\mathbb{E}[C_\infty(x)] < \mathbb{E}[C_0(x)] < \infty$?*

A negative answer to Question 4 would yield a positive answer to Question 5.

References

[1] E. Coffman, P.-J. Courtois, E. Gilbert, and P. Piret, *A distributed clustering process*, J. Appl. Probab., 28 (1991), pp. 737–750. MR1500106

[2] J. van den Berg and R. W. J. Meester, *Stability properties of a flow process in graphs*, Random Structures Algorithms, 2 (1991), pp. 335–341. MR1109700

[3] J. van den Berg and A. Ermakov, *On a distributed clustering process of Coffman, Courtois, Gilbert and Piret*, J. Appl. Probab., 35 (1998), pp. 919–924. MR1671242

[4] M. R. Hilário, O. Louidor, L. T. Rolla, C. M. Newman, S. Sheffield, V. Sidoravicius, *Fixation for distributed clustering processes*, Comm. Pure Appl. Math., 63 7 (2010) pp. 926–934. MR2662427

[5] R. PEMANTLE, *Choosing a spanning tree for the integer lattice uniformly*, Ann. Probab., 19, No. 4 (1991), pp. 1559–1574. MR1127715

[6] K. S. ALEXANDER, *Percolation and minimal spanning forests in infinite graphs*, Ann. Probab., 23, No. 1 (1995), pp. 87–104. MR1330762

[7] R. LYONS, Y. PERES, O. SCHRAMM, *Minimal spanning forests*, Ann. Probab., 34, No. 5 (2006), pp. 1665–1692. MR2271476