Available online at www.sciencedirect.com

ScienceDirect

Discrete Applied Mathematics 155 (2007) 150–160

DISCRETE APPLIED MATHEMATICS

www.elsevier.com/locate/dam

Agent-based randomized broadcasting in large networks☆

Robert Elsässer, Ulf Lorenz, Thomas Sauerwald

Institute for Computer Science, University of Paderborn, Fürstenallee 11, D-33102 Paderborn, Germany

Received 4 October 2004; received in revised form 1 October 2005; accepted 20 April 2006
Available online 11 July 2006

Abstract

Mobile agents are software abstractions that can migrate across the links of a network. They naturally extend the object oriented program style and nicely correspond to agents as examined in game theory. In this paper, we introduce a simple, robust, and efficient randomized broadcast protocol within this mobile agent programming paradigm. We show that by using this scheme, broadcasting enquiries in a random graph of certain density $O(\ln n)$ steps, where n denotes the number of nodes in the graph. Then, we consider bounded degree graphs and prove that we are able to distribute an information among all nodes in $O(D)$ steps, where D denotes the diameter of the graph. We also show that, in contrast to traditional randomized broadcasting (TRB), graphs exist in which agent-based randomized broadcasting requires $\Omega(n^2)$ steps. On the other hand, some graphs which require $\Omega(n \ln n)$ steps to spread the information in the traditional broadcast model, allow very fast agent-based broadcasting. It should be noted that the previously mentioned results are guaranteed with probability $1 - o(1/n)$.

Keywords: Random graphs; Random walks; Mobile agents

1. Introduction

In the agent-based model, as examined in this paper, n agents are randomly distributed in a graph with n nodes. The agents move randomly from one node to another, in rounds, across the edges of the network. At a given time, we inject a piece of information to a node v, which is then called *informed* or *infected*. In the succeeding rounds, informed nodes infect visiting agents and infected agents carry the information to other nodes. The goal is to determine the number of rounds required by the agents to infect all nodes in the network. The agents only know the neighborhood of their hosting nodes; the size of the network, its topology or the 'infected' area are unknown.

The randomized broadcast strategy, as introduced in this paper, is simple, scalable, and robust (i.e., some nodes can be switched off, while their corresponding agents are working in the system). Apart from these characteristics, the algorithm also allows to interleave several tasks. Due to this property, agent-based randomized broadcasting is well suited for asynchronous systems, in which we cannot expect to control the task flow of all participants. The object oriented programming view greatly interferes with this paradigm [13].

This agent-based model is fairly different from the traditional randomized broadcast model (TRB) [9], in which any informed node sends, in each step, some rumor to a randomly chosen neighbor. We will see in Section 4 that there are

☆ This work was partly supported by the IST Program of the EU under contract numbers IST-2001-33116 (FLAGS) and IST-1999-20751 (UpTV). Parts of the results have been presented at MFCS'04 [7].

E-mail address: elsa@upb.de (R. Elsässer).

0166-218X/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.dam.2006.04.035

examples for the agent-based algorithm being asymptotically slower than the corresponding algorithm in the TRB, and vice versa. Nevertheless, we show that for several important graph classes agent-based randomized broadcasting can be performed asymptotically as fast as in the TRB model.

1.1. Specific background

Our main intention is to initiate the analysis of agent-based broadcasting apart from the different application possibilities of this model. However, our attention has been attracted to this model when we studied distributed algorithms in the area of personalized television.

The idea behind personalized television is that a collection of 'personal video recorders' (PVRs) collect information about their users each, and then filter the sky for contributions which each PVR supposes to be interesting for it. The collected information of a PVR is called a user profile. One possibility for a PVR to come to reasonable predictions is to compare its profile with the profiles of other users, and then to base predictions on the opinions of users which have a similar profile. Thereby, two questions arise. First, how to define similarity, but reasonable heuristics exist. Second, how can one find a PVR the users of which are similar to it, i.e., how to find *friends*. The common procedure is to build a database of profiles and compare them centralized [19]. There have also been efforts to work within a distributed database, however, the main focus remains centralistic and it is questionable whether the existing proprietary systems will work in the mass market with more than 10^8 PVRs.

We propose another way to deal with such a large number of PVRs. Assume that every PVR that enters the system, notifies a central instance which places it into a ground network (such as a grid), ensuring that all PVRs are within one connected component. From a practical point of view, this causes no timing problems, since every PVR enters the system only once. Each PVR has the ability to build a specified number of connections to other PVRs, which means that it possesses a couple of half-links at the beginning. A pair of half-links becomes a virtual link across the Internet, as soon as the corresponding PVRs have found each other. In order to achieve this, each PVR is allowed to send out one agent which executes the following protocol: when an agent A enters a node P, then A asks P whether it wants to be a friend of A's origin Q. If P has either free half-links, or the distance between the profile of Q and P is smaller than the distance to one of its neighbors, then P agrees and sends a request directly to Q. If Q wants to be P's friend too, then they establish the link between Q and P. Possibly, P or Q have to resolve an existing link to a processor R. In that case R is notified.

Lots of links are resolved and connected in the system, however, after a finite time the system stabilizes. The reason simply is that there is one cheapest edge in the system. As soon as this edge is detected, it will never be resolved again. Altogether, the procedure leads to a distributed greedy algorithm, and from the edges point of view, the system finishes in a Nash equilibrium. When the system is frozen, the nodes are not interested in changing their connections without further connections being changed simultaneously. The number of steps until the system is frozen depends on the profiles and on the networks which arise during the procedure. Simulations showed that everything stabilizes fast with randomly chosen profiles, however, one subproblem remains: when the algorithm comes close to its final state, a few free half-links occur which should be reconnected as fast as possible. This subproblem can be reduced to the problem of fast broadcasting in the agent-based model.

The system can be seen as a network of n processors, which communicate via a fixed number of agents, ensuring a fast distribution of information, without flooding the network, and without restricting the reach of a message. Since the number of PVRs which enter or possibly leave the system changes slowly, we assume that the number of nodes in our network is fixed. The agent-based view on the given problem fits nicely into known distributed agent protocols, e.g. distance vector routing [14].

1.2. Related work

Mobile agents (MAs) [17] are software abstractions that can migrate across a network, as for example, the Internet. They are the successors of process migration, and therefore in its core as old as distributed computing itself. Several high level programming tools, such as Java or the script language Tcl/tk, support mobility. Aglets, Concordia, Jumping Beans etc. are examples of industrial programming tools which support the MA paradigm. From the software technology's point of view, the MA paradigm serves with benefits like improved locality of reference, the ability to deal with ad hoc ideas as disconnected users, and flexibility. It is worth mentioning that the agent is the basic entity in game theory mostly analyzed within the context of selfish agents [22]. Nevertheless, the MA paradigm stands in concurrency with other techniques like Message Passing. Certainly, many problems addressed by mobility can also be solved even more efficiently and more securely by static clients that exchange messages. We see the difference between MA and Message Passing only in the programmer's point of view. Indeed a mobile agent is a piece of software, and an MA is a data packet which invokes some action on the target computer, from a machine's point of view, MA is very similar to Message Passing. In the same way as a message, the MA must be processed with the help of its recipient. An MA however, is a quite special message, as not every traditional message encodes a piece of a program.

There is also a long history of empirical and theoretical research on *epidemic disease* within cliques and random networks [12,20,1]. Often, mathematical studies about infection propagation make the assumption that an infected person spreads the infection equally likely to any member of the population [18] which leads to a complete graph for the underlying network. Whenever the question is, how fast the disease reaches all persons, the problem reduces to the broadcasting problem in the TRB model. However, in most of these papers, spreaders are only active in a certain time window, and the question of interest is, whether on certain networks modelling personal contacts an epidemic outbreak occurs. Several threshold theorems involving the basic reproduction number, contact number, and the replacement number have been stated. See [12] for a collection of results concerning the mathematics of infectious diseases. Notably, the analysis of epidemic diseases have recently been extended to more generalized random networks with arbitrary degree distributions [20].

Concerning results on the field of *broadcasting*, Frieze and Molloy [10] showed that in a random graph G_p with n vertices, an upper bound of $\Theta(\ln n/n)$ is required on the edge density in order to deterministically broadcast information in $\lceil \log_2 n \rceil$ steps (with high probability). Chen improved this result in [3]. The TRB has also been examined within geometric networks in [16]. It is shown that new information is spread to nodes at distance t with high probability in $O(\ln^{1+\varepsilon} t)$ steps. A similar broadcasting model has been analyzed under the name of rumor-spreading. There, one of n people knows some rumor and any 'knower', in our language an infected person, infects in each round another randomly chosen person of the population. The goal is to determine the number of rounds required for infecting all persons in the system. Pittel [21] proved a nice result, which shows that within $\log_2(n) + \ln(n) + O(1)$ steps they are probably infected. Feige et al. [9] extended the results to different graph classes. Karp et al. [15] showed that, in the so-called random phone call model, the number of messages can be bounded by $O(n \ln \ln n)$.

1.3. Our results

The paper is organized in the following way. We start our analysis by considering randomized agent-based broadcasting in random graphs. We show in Section 2 that within $O(\ln n)$ rounds, every node of a random graph receives the information (with high probability) whenever the generating probability function of the random graph exceeds some certain threshold. In Section 3, we prove that in any bounded degree graph, we can broadcast within $O(D)$ rounds with high probability, where D denotes the diameter of the graph. In Section 4, we consider graph classes on which agent-based broadcasting performs very fast or very slow. We show that while traditional broadcasting can always spread an information within $O(n \ln n)$ steps, graphs exist on which $\Omega(n^2)$ steps are always required in the agent-based model. In contrast to this, on the star we need only $O(\ln n)$ rounds to spread the information (with high probability), while in the TRB model $\Omega(n \ln n)$ steps are needed. We conclude by pointing to some further research directions.

2. Information spreading in random graphs

In this section, we consider the problem of agent-based information spreading in a randomly connected environment. We assume that the underlying network is modelled by a random graph $G_p = (V, E)$ defined as follows: given n and p, generate graph G_p with n vertices by letting each pair be an edge with probability p, independently [2]. Here, we assume that $p = \delta \ln n/n$, where δ is chosen such that the (large) constants $\alpha > 0$ and $\beta > \alpha$ with the following property exist: $\alpha pn \leqslant d_{\min} \leqslant d_{\max} \leqslant \beta pn$, with high probability,[1] where $d_{\min}$ and $d_{\max}$ represent the minimal and maximal vertex degrees in G_p, respectively. The choice of p also implies that the graph is connected with high probability.

As described in the Introduction, for any $t \in \mathbb{N}$, the agents lying at time t on a node $v \in V$ are able to change in round $t+1$ from v to any of its neighbors with probability $1/(d_v + 1)$, where d_v denotes the degree of v. Accordingly,

[1] "With high probability" or "w.h.p." means with probability $1 - o(1/n)$.

an agent remains on its hosting node with the same probability $1/(d_v + 1)$. At some time, a piece of information is placed at one of the nodes, and the goal is to determine the number of rounds needed to distribute this information by n agents in the system. Note that at the beginning, each agent is equally distributed among the nodes of the system.

Let $I(t)$ denote the number of infected nodes at time t. Similarly, $I_a(t)$ denotes the number of infected agents at time t. Let $i(t) = I(t)/n$ and $i_a(t) = I_a(t)/n$ be the fraction of infected nodes and infected agents, respectively, at time t. The healthy nodes and agents are denoted by $H(t) = n - I(t)$ and $H_a(t) = n - I_a(t)$.

We will show that if $p = \delta \ln n/n$ with δ defined above, then every node of G_p is infected after $\mathrm{O}(\log n)$ steps, with high probability. First we analyze the distribution of the agents among the nodes of the graph. If a graph is regular, then each agent jumps to some neighbor of the hosting node with the same probability. We can describe this process by a Markov chain with transition matrix P, where $P_{i,j} = 1/(d+1)$ if $\{i, j\} \in E$ or $i = j$ (d denotes the degree of the nodes in the graph), and $P_{i,j} = 0$ otherwise. Since P is double stochastic, the vector $\mathbf{1} = (1, 1, \dots, 1)$ is an eigenvector of P with eigenvalue 1 and all other eigenvalues are in the range $(-1, 1)$. If we assume that at the beginning each agent is distributed with the same probability among the nodes, then in any step, an arbitrary agent lies on some node of the graph with probability $1/n$. However, if a graph is not regular, then this does not hold. The transition matrix P is then defined by $P_{i,j} = 1/(d_j + 1)$, if $\{i, j\} \in E$ or $i = j$,[2] where d_j describes the degree of node j, and $P_{i,j} = 0$ otherwise. If the vector $(x_1, \dots, x_n)$ denotes the stationary distribution of the ergodic reversible Markov chain described by P, then it holds that $x_i = (d_i + 1)/\left(\sum_{j=1}^{n} d_j + n\right)$ for any $i \in \{1, \dots, n\}$. In the case of a random graph G_p, with p being defined at the beginning of this section, it holds that $x_{\max}/x_{\min} \leqslant (1 + \beta)/(1 + \alpha)$, where α and β are the constants defined in the first paragraph of this section.

Concerning the expansion properties of G_p, [2,5] yield the following theorem.

Theorem 1. *Let $G_p = (V, E)$ be a random graph with n nodes, generated by $p = \delta \ln n/n$. Then, it holds with high probability that the edge expansion of the graph, i.e., $\min\{|E(X, \overline{X})|/m \mid X \subset V$ and $|X| = m\} \geqslant \gamma pn$ for any $m \leqslant n/2$, where $E(X, \overline{X})$ represents the set of edges between X and $\overline{X}$, and $\gamma > 0$ is a constant value.*

This theorem implies that, with high probability, the distribution of the agents among the nodes of G_p becomes very close to the stationary distribution within $\mathrm{O}(\ln n)$ steps [5,23,8].

We assume in the following lemmas that a constant τ exists such that, with probability at least $1 - 1/e^{\tau}$, for some arbitrary node $v \in V$ there is at least one agent on v. First we analyze the case when $I(t) \leqslant q \ln \ln n$, where q is a properly chosen large constant value.

Remark 2. If $I(t) \leqslant q \ln \ln n$ at some time t and $I_a(t) \geqslant 1$, then in step $t + 1$ another node will be infected with probability $1 - \mathrm{O}(\ln \ln n/ \ln n)$, where q is a properly chosen large constant.

Proof. Obviously, each node has a degree of at least αpn (with high probability). In step $t+1$, the infected agent jumps from an infected node v to some infected vertex with a probability of at most $q \ln \ln n/(\alpha pn)$. Since $p \geqslant \delta \ln n/n$, the remark follows. □

Now consider the case when $q \ln \ln n \leqslant I(t) \leqslant q \ln n$.

Lemma 3. *Let S be the set of infected nodes in G_p at time t, and assume that each node $i \in S$ has at most $\max\{c_\delta \ln |S|, c_\delta p|S|\}$ neighbors in S, where c_δ is a constant value. If $I(t) \in [q \ln \ln n, q \ln n]$, where $I(t) = |S|$, then a constant c exists so that $I(t + 1) \geqslant I(t)(1 + 1/c)$ with probability $1 - \mathrm{o}(1/ \ln n)$.*

Proof. Let $p_1 = 1/e^{\tau}$ be the probability that an arbitrary node is empty (i.e., no agent is lying on it). Then, a second node is empty with the conditional probability p_2 (given that a node exists, which is already empty), where $p_2 < p_1$. Using the Chernoff bound on the tail of a binomial distribution [4,11], the probability for having more than $(2/e)^{\tau} I(t)$

[2] Letting $P_{i,i}$ be nonzero is a simple device to avoid periodicity problems.

infected empty nodes is

$$P_{\text{empty}} < \sum_{i=(2/e)^{\tau} I(t)}^{I(t)} \binom{I(t)}{i} \left(\frac{1}{e^{\tau}}\right)^{i} \left(1 - \frac{1}{e^{\tau}}\right)^{I(t)-i} \tag{1}$$

$$\leqslant \left(\frac{1}{2}\right)^{(2/e)^{\tau} I(t)} \left(\frac{e^{\tau}-1}{e^{\tau}-2^{\tau}}\right)^{I(t)(1-(2/e)^{\tau})} = \mathrm{o}\left(\frac{1}{\ln n}\right), \tag{2}$$

whenever q is large enough. Now we show that among these $|R| = (2/e)^{\tau} I(t)$ nonempty nodes, at least $|R|/2$ will propagate the information to some healthy nodes in step $t+1$ with probability $1 - \mathrm{o}(1/\ln n)$. Let $d_S(u)$ denote the number of neighbors of a node $u \in V$ in the set $S \subset V$ of infected nodes. From more than $|R|/2$ nonempty infected nodes, the agents do not jump to healthy nodes with probability

$$\begin{aligned} P_{\text{fail}} &\leqslant \sum_{i=|R|/2}^{|R|} \binom{|R|}{i} \left(\frac{\max_{u\in S} d_s(u)}{\alpha p n}\right)^{i} \left(1 - \frac{\max_{u\in S} d_s(u)}{\alpha p n}\right)^{|R|-i} \\ &\leqslant \sum_{i=|R|/2}^{|R|} \binom{|R|}{i} \left(\frac{c_{\delta} \ln \ln n}{\alpha \ln n}\right)^{i} \left(1 - \frac{c_{\delta} \ln \ln n}{\alpha \ln n}\right)^{|R|-i} \\ &\leqslant \left(\frac{2c_{\delta} \ln \ln n}{\alpha \ln n}\right)^{|R|/2} \left(2\left(1 - \frac{c_{\delta} \ln \ln n}{\alpha \ln n}\right)\right)^{|R|/2} = \mathrm{o}\left(\frac{1}{\ln n}\right). \end{aligned}$$

In order to obtain the first and second inequality, we use the observation that P_{fail} is maximized when p and d_u are minimized and the inner degree $d_S(u)$ is maximized for all $u \in S$.

Summarizing, at least $|R|/2$ infected agents jump to uninfected nodes in step $t+1$ with probability $1 - \mathrm{o}(1/\ln n)$. On the other hand, applying the Chernoff bound we obtain that after the $(t+1)$st step, any newly infected node has less than 3 previously infected agents (w.h.p.). □

After $I(t) \geqslant q \ln n$ is achieved, similar methods to that of [9] can be used to show fast broadcasting. However, in this agent-based model we can only guarantee that at least a constant fraction of the infected nodes are able to forward the information. We have to show that another constant fraction of the agents positioned on these nodes infect some healthy nodes in the next step. We consider now the case when $q \ln n \leqslant I(t) \leqslant \sqrt[4]{n}$.

Lemma 4. *Let S be the subset of infected nodes at some time t with $I(t) = |S|$. If $I(t) \in [q \ln n, \sqrt[4]{n}]$, then it holds that $I(t+1) \geqslant I(t)(1 + 1/c)$ with probability $1 - \mathrm{o}(1/n)$, where c is a proper constant value.*

Proof. We assumed that any node in S is empty with probability at most $1/e^{\tau}$ (cf. Lemma 3). Due to Theorem 1, we know that at least $\gamma |S| p n$ edges are connecting S and $V \backslash S$. We may assume that every node in S has degree at most $\beta p n$. Therefore, any agent lying on some infected node $u \in S$ jumps to an uninfected node with probability at least $d_S(u)/(\beta p n)$, where $d_S(u)$ represents the number of u's neighbors in S. Therefore, the expected number of infected agents jumping to uninfected nodes is at least

$$\sum_{u\in S} (1 - 1/e^{\tau}) d_S(u)/(\beta p n) = (1 - 1/e^{\tau}) \gamma S/\beta = \Omega(S). \tag{3}$$

Since $I(t) \geqslant q \ln n$ applying the Chernoff bounds [11] we conclude that a constant c'' exists so that from each of at least $I(t)/c''$ nodes an agent will jump to healthy nodes with probability $1 - \mathrm{o}(1/n^2)$, whenever q is large enough. The probability that j of these infected agents share the same node after the $(t+1)$st step is smaller than $\binom{I_a(t)}{j}(1/n)^j = \mathrm{O}(1/n^{3j/4})$, where $j \in \mathbb{N}$. Thus, with probability $1 - \mathrm{o}(1/n^2)$, each newly infected node has at most 2 previously infected agents lying on it. Since $I(t)/c''$ infected agents are jumping to healthy nodes in step $t+1$ (with probability $1 - \mathrm{o}(1/n^2)$), and each newly infected node has at most 2 of such agents, it follows that at least $I(t)/(2c'')$ nodes become infected in step $t+1$ (with probability $1 - \mathrm{o}(1/n^2)$). □

Lemma 5. *Let S be the subset of infected nodes at some time t with $I(t)=|S|$. If $\sqrt[4]{n}\leqslant I(t)\leqslant n-\sqrt[4]{n}$, then it holds that $I(t+1)\geqslant I(t)(1+H(t)/(cn))$, for some constant c.*

Proof. First, we determine the number of infected agents at time t. Let $I(t)<n/2$. We assumed that a constant τ exists so that an arbitrary node is empty with probability $1/e^{\tau}$. Again, we apply Eq. (3) concluding that $\Omega(S)$ agents jump to uninfected nodes with probability $1-o(1/n^2)$.

Similarly, if $H(t)\leqslant n/2$, then a constant c' exists so that at least $H(t)/(2c')$ agents infect some uninfected nodes. Summarizing, a constant c exists so that at least $4I(t)H(t)/(cn)$ agents propagate the infection to uninfected nodes in step $t+1$ (with probability $1-o(1/n^2)$), where $c>4$.

Ignoring the probability for which less than $4I(t)H(t)/(cn)$ infected agents jump to uninfected nodes, an arbitrary healthy node remains uninfected with probability

$$p_1\leqslant\left(1-\frac{1}{H(t)}\right)^{4I(t)H(t)/(cn)}\leqslant(1/e)^{\frac{4I(t)}{cn}}.$$

The probability for a second node remaining uninfected is $p_2\leqslant p_1$. Similarly, a node i remains uninfected with the conditional probability p_i (given that $i-1$ nodes are uninfected), where it holds that $p_i\leqslant p_1$. Therefore, more than $H(t)-I(t)H(t)/(cn)$ nodes remain uninfected after the $(t+1)$st step with probability

$$P(t)\leqslant\sum_{i=H(t)-\frac{I(t)H(t)}{cn}}^{H(t)}\binom{H(t)}{i}p_1^i(1-p_1)^{H(t)-i}.$$

Noting that $p_1\leqslant(1/e)^{4I(t)/(cn)}$ and that $P(t)$ is minimized if p_1 is maximized, we get

$$P(t)\leqslant\sum_{i=H(t)-\frac{I(t)H(t)}{cn}}^{H(t)}\binom{H(t)}{i}\left(\frac{1}{e}\right)^{4i\cdot I(t)/(cn)}\left(1-\left(\frac{1}{e}\right)^{4I(t)/(cn)}\right)^{H(t)-i}.$$

Since $(H(t)-I(t)H(t)/cn)/H(t)>(1/e)^{4I(t)/(cn)}$, using the results of e.g. [11] we obtain

$$P(t)\leqslant\left(\frac{H(t)}{H(t)-\frac{I(t)H(t)}{cn}}\right)^{H(t)-\frac{I(t)H(t)}{cn}}\left(\frac{H(t)}{\frac{I(t)H(t)}{cn}}\right)^{\frac{I(t)H(t)}{cn}}\left(\frac{1}{e}\right)^{\frac{4I(t)}{cn}(H(t)-\frac{I(t)H(t)}{cn})}\left(1-\left(\frac{1}{e}\right)^{\frac{4I(t)}{cn}}\right)^{\frac{I(t)H(t)}{cn}}.$$

Then, it holds that

$$\begin{aligned}
P(t)&\leqslant\left(\frac{1}{1-\frac{I(t)}{cn}}\right)^{H(t)(1-\frac{I(t)}{cn})}\left(\frac{cn}{I(t)}\right)^{\frac{I(t)H(t)}{cn}}\left(\frac{1}{e}\right)^{\frac{4I(t)H(t)}{cn}(1-\frac{I(t)}{cn})}\left(1-\left(\frac{1}{e}\right)^{\frac{4I(t)}{cn}}\right)^{\frac{I(t)H(t)}{cn}}\\
&\leqslant\left(\frac{1}{1-\frac{I(t)}{cn}}\right)^{H(t)(1-\frac{I(t)}{cn})}\left(\frac{cn}{I(t)}\right)^{\frac{I(t)H(t)}{cn}}\left(\frac{1}{e}\right)^{\frac{4I(t)H(t)}{cn}(1-\frac{I(t)}{cn})}\left(1-\left(1-\frac{4I(t)}{cn}\right)\right)^{\frac{I(t)H(t)}{cn}}\\
&\leqslant\left(\frac{1}{1-\frac{I(t)}{cn}}\right)^{H(t)(1-\frac{I(t)}{cn})}\left(\frac{1}{e}\right)^{\frac{4I(t)H(t)}{cn}(1-\frac{I(t)}{cn})}4^{\frac{I(t)H(t)}{cn}}\\
&\leqslant e^{\frac{I(t)}{cn-I(t)}H(t)(1-\frac{I(t)}{cn})}\left(\frac{1}{e}\right)^{\frac{4I(t)H(t)}{cn}(1-\frac{I(t)}{cn})}4^{\frac{I(t)H(t)}{cn}}\\
&\leqslant\left(e\left(\frac{1}{e}\right)^{4(1-\frac{I(t)}{cn})}4\right)^{\frac{I(t)H(t)}{cn}}\leqslant\left(\frac{2}{e}\right)^{\frac{2\sqrt[4]{n}}{c}}
\end{aligned}$$

and the lemma follows. □

If $H(t)\leqslant\sqrt[4]{n}$, then we can show the following lemma.

Lemma 6. *Let S be a set of healthy nodes in G_p. If $H(t) \leqslant \sqrt[4]{n}$ at some time t, where $H(t) = |S|$, then a constant $c > 0$ exists so that an arbitrary node $v \in S$ is infected in step $t+1$ with probability $1/c$. Moreover, after $\mathrm{O}(\ln n)$ rounds, all nodes of the graph are infected with high probability.*

Proof. Applying Eq. (1) with $I(t) > n - \sqrt[4]{n}$, it follows that, with probability $1 - \mathrm{o}(1/n^2)$, in step t at least $n(1-(2/e)^\tau)$ infected nodes are not empty. Due to the distribution of the edges out of S, $\lceil |S|/2 \rceil$ nodes have at least pn/c' neighbors within this set of $n(1-(2/e)^\tau)$ infected nonempty nodes (with probability $1 - \mathrm{o}(1/n^2)$), where c' is a constant. Since each node v belongs to this set of $|S|/2$ nodes with the same probability, v becomes infected in the next step with probability $1/c$, where c is a proper constant value.

Whenever $H(t) \geqslant q \ln n$, there exists a constant c'' such that $H(t+1) \leqslant H(t)(1 - 1/c'')$ with probability $1 - \mathrm{o}(1/n^2)$. In order to see this, let P_l be the probability that more than $H(t)/c''$ nodes remain uninformed after the tth step. Now, when a node remains uninformed, then a second node is also uninformed with some conditional probability less than $1/c$, and P_l fulfils the inequality

$$P_l \leqslant \sum_{i=|S|/c''}^{|S|/2} \binom{|S|/2}{i} (1/c)^i (1 - 1/c)^{|S|/2-i} \leqslant \mathrm{o}(1/n^2) \tag{4}$$

for some constant c'' chosen properly. When $q \ln \ln n \leqslant H(t) \leqslant q \ln n$, then inequality (4) implies that a constant c'' exists so that $H(t+1) \leqslant H(t)(1 - 1/c'')$ with probability $1 - \mathrm{o}(1/\ln n)$. This implies that within $\mathrm{O}(\ln n)$ rounds, the number of healthy nodes is reduced to $q \ln \ln n$. Finally, since each remaining node is infected with probability at least $1/c$, within $\mathrm{O}(\ln n)$ additional steps every node of S is infected with probability $1 - \mathrm{o}(1/n)$. □

We are now ready to state the following theorem concerning the spread of an information in random graphs using the agent-based model.

Theorem 7. *Let $G_p = (V, E)$ be a random graph with n vertices, where $p = \delta \ln n/n$. At the beginning, we distribute somehow n agents among the nodes of G_p, and allow each agent, which lies on some node $v \in V$, to change with probability $1/(d_v + 1)$ to one of the neighbors of v. Assuming that at some time an arbitrary node receives an information, after $\mathrm{O}(\log n)$ rounds the information is completely distributed in the whole system with high probability.*

Proof. After a piece of information is injected to some node $v \in V$, then within $\mathrm{O}(\ln n)$ rounds the distribution of the agents becomes very close to the stationary distribution of the Markov chain described by the matrix P. Within $\mathrm{O}(\ln n)$ additional rounds one of the agents will visit v, with high probability, and carry the information further.

Now, we consider the following algorithm. Until $I(t) \leqslant q \ln \ln n$, we allow only this one agent to transmit the infection from one node to another. Due to Remark 2, within $\mathrm{O}(\ln n)$ steps we achieve $I(t) \geqslant q \ln \ln n$ (w.h.p.). In the following rounds, we call a step successful, if within this step at least $I(t)H(t)/(cn)$ uninfected nodes become infected, where c is the constant defined in Lemmas 3–5. Otherwise, the step is called unsuccessful and the agents that have jumped to healthy nodes become healthy (instead of infecting the nodes). In each successfull step, we choose $I(t)H(t)/(cn)$ newly infected nodes, uniformly at random, among all newly infected nodes, and heal all other nodes which became infected in this last step. Clearly, each node gets the information from an informed agent. The distribution of the other agents, not seen before to infect some node, differ only in a constant factor from the stationary distribution. Since a constant fraction of the agents infecting some node in an arbitrary step carry the information further to uninfected nodes in the next step, there exists a constant τ such that in any step, each infected node is empty with probability at most $1/e^\tau$.

Since the assumptions of Lemmas 3–5 are fulfilled, we need $\mathrm{O}(\ln n)$ steps to achieve $H(t) \leqslant \sqrt[4]{n}$ (w.h.p.). Finally, if $H(t) \leqslant \sqrt[4]{n}$, we do not speak anymore about successful or unsuccessful rounds and allow each infected agent to infect any healthy node without becoming itself healthy at all. Using Lemma 6, we can show that with high probability, within another $\mathrm{O}(\ln n)$ steps all nodes in the graph become infected.

Due to the arguments described above, the algorithm described in this proof spreads the information among all nodes in the graph within $\mathrm{O}(\ln n)$ steps. Obviously, the algorithm presented at the beginning of this section performs faster than the one described here, and therefore the theorem holds. □

The results of Theorem 7 can easily be generalized to the other traditional random graph model [2]: given n and m, let each graph with n vertices and m edges occur with probability $\binom{N}{m}^{-1}$, where $N = \binom{n}{2}$. The random variable $G_{n,m}$ represents a graph generated in this way. If $m = \delta n \ln n$ with δ being large enough, then the results described in this section also hold for $G_{n,m}$.

3. Information spreading in bounded degree graphs

In the previous section, we determined the number of steps required to distribute the information in a system described by a random graph. In this section, we consider the previously mentioned distribution problem in a more restricted environment. We also slightly modify the randomized scheme in the following way. An agent lying at time t on some node $v \in V$ can jump to one of the neighbors of v with probability $1/(2d_v)$, where d_v denotes the degree of node v. Accordingly, the agent remains on v with probability $\frac{1}{2}$. We assume throughout this section that d_v is bounded by some constant value c for any node v.

We show that with high probability, in at most $O(D)$ steps the information is distributed among the nodes of the system, where D denotes the diameter of G. First we analyze the probability distribution of the agents among the nodes of the graph. As described in the previous section, each agent is equally distributed at the beginning. Then, the following lemma can be stated.

Lemma 8. *Let $G = (V, E)$ be a graph and let P be a transition matrix defined in the following way: $P_{i,i} = \frac{1}{2}$ for any $i \in \{1, \ldots, n\}$, $P_{i,j} = 1/(2d_j)$ if $\{i, j\} \in E$, where d_j describes the degree of node j, and $P_{i,j} = 0$ otherwise. We assume that a constant c exists so that $c \geqslant d_i$ for any $i \in \{1, \ldots, n\}$. Let $x = (x_1, \ldots, x_n)$ be defined by $x^{\mathrm{T}} = P^t \cdot (1/n)\mathbf{1}^{\mathrm{T}}$, where t is an arbitrary chosen integer. Then, it holds that $x_i \geqslant 1/(cn)$ for any $i \in \{1, \ldots, n\}$.*

Proof. The matrix P represents the transition matrix of an ergodic reversible Markov chain with stationary distribution π, where $\pi_i = d_i / \sum_{j_1}^{n} d_j$. Vector π is an eigenvector of P with eigenvalue 1. Due to the properties of P, its eigenvalues are real and they lie in the range $[0, 1]$. On the other hand, the matrix P can be viewed as a diffusion matrix of a node-weighted graph with the structure of G and node-weights d_i [8]. Then, $P = I - \frac{1}{2} L\mathbf{D}^{-1}$, where I is the identity matrix, L is the Laplacian of G [6], and $\mathbf{D}$ is the diagonal matrix containing d_i in the ith diagonal entry. The initial load distribution has the form $w^0 = (1/n)(1, \ldots, I)$, and it holds that $w^{k+1} = P w^k$ in every iteration step k, where w^k denotes the load distribution in G after k steps. Then, it also holds that

$$w_i^{k+1} = w_i^k - \sum_{j \in N(i)} \frac{1}{2} \left(\frac{w_i^k}{d_i} - \frac{w_j^k}{d_j} \right),$$

where $N(i)$ defines the set of neighbors of node i. We show that $\min_{i=1}^{n} w_i^k / d_i \leqslant \min_{i=1}^{n} w_i^{k+1} / d_i$ for any k. Assume first that $w_l^k / d_l = \min_{i=1}^{n} w_i^k / d_i$ for some $l \in \{1, \ldots, n\}$. For any node $i \in V$ it holds that

$$w_i^{k+1} = w_i^k - \sum_{j \in N(i)} \frac{1}{2} \left(\frac{w_i^k}{d_i} - \frac{w_j^k}{d_j} \right) = \frac{w_i^k}{2} + \frac{1}{2} \sum_{j \in N(i)} \frac{w_j^k}{d_j} \geqslant \frac{w_i^k}{2} + \frac{d_i}{2} \cdot \frac{w_l^k}{d_l}.$$

Thus,

$$\frac{w_i^{k+1}}{d_i} \geqslant \frac{1}{2} \left(\frac{w_i^k}{d_i} + \frac{w_l^k}{d_l} \right) \geqslant \frac{w_l^k}{d_l}.$$

Since $\min_{i=1}^{n} w_i^0 / d_i \geqslant 1/cn$, the lemma holds. □

We are now ready to state the following theorem.

Theorem 9. *Let $G = (V, E)$ be a graph, where $d_i \leqslant c$, c being constant, for any $i \in V$. We assume that at the beginning, n agents are equally distributed among the nodes of G, and at some time an information is placed on a node of G.*

If the movements of the agents satisfy the transition probabilities described by the transition matrix P *(as defined in Lemma* 8), *then within* $\mathrm{O}(D)$ *steps every node will get the information by some agent with probability* $1 - \mathrm{o}(1/n)$.

Proof. Due to Lemma 8, any agent lies in each step on some arbitrary node i with a probability of at least $1/(cn)$. Since the diameter of a graph with bounded vertex degree is $\Omega(\log n)$, the information fails to traverse a shortest path between two vertices in G in less than $6c^2(D + 3\ln n)$ steps with probability $\mathrm{o}(1/n^2)$. Hence, the information reaches all vertices of the graph within $\mathrm{O}(D)$ steps with probability $1 - \mathrm{o}(1/n)$. □

4. Best-case and worst-case graphs

In this section, we consider graph classes on which we can broadcast in the agent-based model very fast or very slowly (with high probability). The agents lying on the nodes of some graph $G = (V, E)$ are able to change from a node $v \in V$ to any of its neighbors with probability $1/(d_v + 1)$, where d_v denotes the degree of i. Accordingly, an agent remains on its hosting node with the same probability $1/(d_v + 1)$. Here, we assume that the agents are lying on the nodes according to the stationary distribution of the Markov chain determined by the transition matrix P as defined in Section 2.

It is known that in the TRB model at most $\mathrm{O}(n \ln n)$ steps are needed in order to broadcast an information among all nodes of an arbitrary graph. We show in the next theorem that this bound cannot be achieved in the agent-based model.

Theorem 10. *There exist graphs* $G = (V, E)$ *in which* $\Omega(n^2)$ *rounds are necessary for broadcasting in the agent-based model.*

Proof. Consider the graph $G = (V, E)$ with n vertices constructed as follows: the first $n/2$ vertices, labelled $1, \ldots, n/2$, are connected with each other mutually, forming a complete graph with $n/2$ vertices. Vertex $n/2$ is additionally connected to vertex $n/2+1$. Then, we connect the vertices $n/2 + j$ and $n/2 + j + 1$ for any $j \in \{1, \ldots, n/2 - 1\}$, by letting the last $n/2$ nodes form a path of length $n/2$. An agent is lying on node $i \in \{1, \ldots, n/2 - 1\}$ with probability $n/(n^2/2 + n + 3)$. Node $n/2$ has an agent with probability $(n + 2)/(n^2/2 + n + 3)$. All other nodes excepting n have an agent with probability $3/(n^2/2 + n + 3)$, and finally on n a certain agent is lying with probability $2/(n^2/2 + n + 3)$.

With some constant probability, each agent lies on some node $i \in \{1, \ldots, n/2\}$. Letting the information be placed somewhere, node n receives the information only if one of the agents traverses the whole path from node $n/2 + 1$ to node n. Using simple probability theory, an agent traverses the path (without dropping first into the complete part) with probability $1/n$. Noting that an agent jumps from node $n/2$ to node $n/2 + 1$ with probability $\mathrm{O}(1/n)$, we obtain the statement of the theorem. □

Now we study a graph on which an exponential gap between the speed of information spreading in the traditional broadcast model and in the agent-based model occurs. Let $G_u = (V, E)$ be a graph, which consists of $2k$ levels, where each level $i \in \{0, \ldots, 2k - 1\}$ contains $2^{(i \bmod k)}$ vertices. We connect the vertices of two consecutive levels i and $i + 1$ mutually, obtaining for any $(i \bmod k) \in \{0, \ldots, k - 2\}$ a complete bipartite graph between levels i and $i + 1$. Additionally, we connect the nodes between levels $k - 1$ and $2k - 1$ in the same way. Using the techniques described in the proof of Theorem 10, it can be shown that we need $\Omega(k2^k)$ steps to propagate the information through the network. In the TRB model, only $\mathrm{O}(k)$ steps are required. Thus, there is an exponential gap between the speed of broadcasting in these two models.

In contrast to the previous paragraph, the star is a good-natured graph for broadcasting in our agent-based model. Despite the fact that the star has only $\Theta(n)$ edges, whereas the complete graph has $\Theta(n^2)$, the time of a broadcast is $\mathrm{O}(\ln n)$ with high probability for both. Surprisingly, the infection among the agents is even faster than on the complete graph. The reason for this is the outstanding position of the central node, which we call z, and we denote the number of agents visiting z at time t by $z(t)$.

Theorem 11. *Let* $\mathbf{S}_n = (V, E)$ *be the star with n vertices and let* z *be its central node. In* $\mathbf{S}_n$ *we equally distribute n agents among the nodes and allow each agent lying on some node* $v \in V\backslash\{z\}$ *to change with probability* $\frac{1}{2}$ *to* z. *Any agent lying on* z *is allowed to change with probability* $p_v = 1/n$ *to some noncentral node* v. *Assuming that at some time an arbitrary node receives an information, after* $\mathrm{O}(\ln n)$ *rounds the information is completely distributed in the whole system with probability* $1 - \mathrm{o}(1/n)$.

Proof. First, we show that the node z is infected in constant expected time. If $i_a(t) > 0$, then we may suppose that the infected agent is on a noncentral node $v \in V\backslash\{z\}$. Since the agent jumps to z with probability $\frac{1}{2}$, the agent infects z after 2 expected rounds. If $i_a(t) = 0$ and $i(t) > 0$, then let $v \in V\backslash\{z\}$ be an infected noncentral node. After an agent reaches node v, we are able to apply the previous case. The probability that a certain uninfected agent A reaches node v within two rounds is at least $1/(2n)$. To see this, first consider the case when A is on another noncentral node v'. The agent leaves v' with probability $\frac{1}{2}$. However, if A is on z, it jumps to v with probability $1/n$. Thus, the probability that at least one agent reaches v within two rounds is: $1 - (1 - 1/(2n))^n \geqslant 1 - e^{-1/2}$. Therefore, the expected time until an agent jumps to v is $2/(1 - e^{-1/2}) \leqslant 2 \cdot 3 = 6$. Applying the Chernoff bound [11], node z is infected within $\mathrm{O}(\ln n)$ steps with probability $1 - \mathrm{o}(1/n)$.

Now we will concentrate on the infection among the agents. First we show that $\max\{z(t), z(t+1)\} \geqslant n/5$ for any $t \in \mathbb{N}$ with probability $1 - \mathrm{o}(1/n^2)$. To prove this, we may assume $z(t) < n/5$. Then, at least $4n/5$ agents are on $V\backslash\{z\}$ at time t. Since each of these agents jumps to z with probability $\frac{1}{2}$, we can apply the Chernoff bound [4,11] to show that $z(t+1) \geqslant n/5$ with probability $1 - \mathrm{o}(1/n^2)$. As a consequence after two rounds, at least $n/5$ agents become infected with a probability of $1 - \mathrm{o}(1/n)$.

Now we turn our attention to the infection of nodes. The probability that an uninfected node remains uninfected after t rounds is $(((n-1)/n)^{n/5})^t \leqslant (1/e)^{t/5}$. Hence, for some $t = \Theta(\ln n)$, we obtain the theorem. □

We have seen that agent-based broadcasting can be performed very fast on a star. Let us now consider the graph G_s which is very similar to the star and is defined in the following way: the first $n/2$ nodes of G_s, labelled $0, 1, \ldots, n/2-1$, of G_s are connected mutually with each other, forming a complete graph with $n/2$ nodes, while a node $n/2 + j - 1$ is only connected to $j - 1$. Using the techniques of Theorem 11, it can be shown that agent-based broadcasting requires $\Omega(n \ln n)$ rounds. This example shows that, although the diameter of G_s is $\mathrm{O}(1)$ and the graph is very similar to the star, agent-based randomized broadcasting performs slowly.

5. Conclusion

In this paper, we analyzed the performance of randomized broadcasting within the agent-based model. We have shown that with high probability, we can distribute within $\mathrm{O}(\ln n)$ steps an information among the nodes of a random graph of certain density. We proved that broadcasting in a bounded degree graph can be performed with high probability in $\mathrm{O}(D)$ steps, where D represents the diameter of the graph. We considered examples of graphs, in which agent-based broadcasting is very fast or very slow. We also pointed out some differences between broadcasting in the agent-based model and the TRB model.

Our main intention was to initiate the analysis of broadcasting in the agent-based model. However, the results of this paper can only be viewed as a first step in this direction, and there are still several interesting open problems which are worth to be analyzed.

In our model we assumed that the number of agents equals the number of nodes. In the case when the number of agents is a constant fraction of the size of the network, then the same asymptotic results also hold. However, it would be interesting to know how the runtime changes when the number of agents (and the overall communication complexity within one step) decreases significantly. Another question is how the runtime depends on the initial distribution of the agents in worst and best case graphs. It is also an open problem how gossiping performs in this agent-based model.

Acknowledgment

We are grateful to the referees for helping us to improve the quality of the paper.

References

[1] F. Ball, A threshold theorem for the reed-frost chain-binomial epidemic, J. Appl. Probab. 20 (1983) 153–157.
[2] B. Bollobas, Random Graphs, Academic Press, New York, 1985.
[3] H. Chen, Threshold of broadcast in random graphs, DIMACS Technical Report 97-12, 1997.
[4] H. Chernoff, A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, Ann. Math. Statist. 23 (1952) 493–507.

[5] F.R.K. Chung, Spectral Graph Theory, CBMS Regional Conference Series in Mathematics, vol. 92, American Mathematical Society, Providence, RI, 1997.
[6] D.M. Cvetkovic, M. Doob, H. Sachs, Spectra of Graphs, third ed., Johann Ambrosius Barth, 1995.
[7] R. Elsässer, U. Lorenz, T. Sauerwald, Agent based information handling in large networks, in: Proceedings of MFCS'04, 2004, pp. 686–698.
[8] R. Elsässer, B. Monien, R. Preis, Diffusion schemes for load balancing on heterogeneous networks, Theory Comput. Systems 35 (2002) 305–320.
[9] U. Feige, D. Peleg, P. Raghavan, E. Upfal, Randomized broadcast in networks, Random Struct. Algorithm 1 (4) (1990) 447–460.
[10] A.M. Frieze, M. Molloy, Broadcasting in random graphs, Discrete Appl. Math. 54 (1994) 77–79.
[11] T. Hagerup, C. Rüb, A guided tour of Chernoff bounds, Inform. Process. Lett. 36 (6) (1990) 305–308.
[12] H.W. Hethcore, Mathematics of infectious diseases, SIAM Rev. 42 (2000) 599–653.
[13] J. Hylton, K. Manheimer, F.L. Drake, B. Warsaw, R. Masse, G. van Rossum, Knowbot programming: system support for mobile agents, in: Proceedings of IWOOOS'96, 1996, pp. 8–13.
[14] A.A. Kaizar, J.T. Mayes, A.R. Mikler, Agent-based distance vector routing, Proceedings of the Third International Workshop MATA, 2001, pp. 41–50.
[15] R. Karp, C. Schindelhauer, S. Shenker, B. Vöcking, Randomized rumor spreading, in: Proceedings of FOCS'00, 2000, pp. 565–574.
[16] D. Kempe, J. Kleinberg, A. Demers, Spatial gossip and resource location protocols, in: Proceedings of STOC'01, 2001, pp. 163–172.
[17] D. Kotz, F. Mattern, Agent systems, mobile agents, and applications, Proceedings of Second International Symposium on Agent Systems and Applications and Fourth International Symposium on Mobile Agents, ASA/MA, 2000.
[18] A.G. McKendrick, W.O. Kermack, Contributions to the mathematical theory of epidemics, in: Proceedings of the Royal Society, 1927, pp. 700–721.
[19] A. Nakamura, N. Abe, H. Matoba, K. Ochiai, Automatic recording agent for digital video server, in: Proceedings of the Eighth ACM International Conference on Multimedia, 2000, pp. 57–66.
[20] M.E.J. Newman, The spread of epidemic disease on networks, Phys. Rev. E 66 (2002) 016128.
[21] B. Pittel, On spreading rumor, SIAM J. Appl. Math. 47 (1) (1987) 213–223.
[22] T. Roughgarden, E. Tardos, How bad is selfish routing?, in: Proceedings of FOCS'00, 2000, pp. 93–102.
[23] A. Sinclair, M. Jerrum, Approximate counting, uniform generation and rapidly mixing Markov chains, Inform. and Comput. 82 (1989) 93–133.