Exploring Complex Networks with Failure-prone Agents

Arles Rodríguez[1,2], Jonatan Gómez[1], Ada Diaconescu[3]

[1]ALIFE Research Group, Universidad Nacional de Colombia
[2]Fundación Universitaria Konrad Lorenz
[3]Telecom ParisTech, LTCI CNRS Paris, France
arlese.rodriguezp@konradlorenz.edu.co, jgomezpe@unal.edu.co, ada.diaconescu@telecom-paristech.fr

Abstract. Distributed data-collection and synchronization is essential in sensor networks and the Internet of Things (IoT), as well as for data-replication in server farms, clusters and clouds. Generally, such systems consist of a set of interconnected components, which cooperate and coordinate to achieve a collective task, while acting locally and being failure-prone. An important challenge is hence to define efficient and robust algorithms for data collection and synchronisation in large-scale, distributed and failure-prone platforms. This paper studies the performance and robustness of different multi-agent algorithms in complex networks with different topologies (Lattice, Small-world, Community and Scale-free) and different agent failure rates. Agents proceed from random locations and explore the network to collect local data hosted in each node. Their exploration algorithm determines how fast they cover unexplored nodes to collect new data, and how often they meet other agents to exchange complementary data and speed-up the process. Two exploration algorithms are studied: one random and one using a stigmergy model (that we propose). Experimental results show how network topologies and agent failure-rates impact data-collection and synchronization, and how a stigmergy-based approach can improve performance and success rates across most scenarios. We believe these results offer key insights into the suitability of various decentralised algorithms in different networked environments, which are increasingly at the core of modern information and communication technology (ICT) systems.

1 Introduction

Sensor networks, server farms and clouds consist of numerous components (e.g. servers, processes, robots) interconnected via (complex) networks. They cooperate and coordinate their actions towards some overall objective, may share common resources, and appear to an end-user as a single system [1]. An important field of study here relates to how fast distributed processes, interconnected via such complex networks, can achieve collective objectives (e.g. data collection, synchronisation or processing); and how the particular topological properties of complex networks impact such performance [2–5]. These aspects are key in secure communications [4], logging and machine replication in databases [6]; and, information-processing and consensus-making in sensor networks [7]. A particular challenge here represents data collection from a network's components, both as a stand-alone objective, e.g., in sensor networks, or as an underlying task for data synchronisation. This paper focuses on decentralised data collection in complex networks. Important challenges must be addressed here, as networked components can only act locally and may fail unexpectedly.

Previous works [8,9] have studied data-collection techniques based on failure-prone agents. Analysed approaches included Random walks, Lévy walks and Stigmergy. Agents explored a targeted space based on selected algorithms, in order to collect local information and share it with other agents (that they could meet) [8]. [9] presents agents that collect information from distributed sources and can fail and/or provide unreliable information defining collective information as agregation of information that agents collect individually. [10] showed how data-collection can be speeded-up by algorithms that favour exploration of new paths and the exchange of new information with other agents. It also showed how mechanisms that favour exploration and achieve faster data-collection are more resistant to failure than those that focus on increasing inter-agent communication. Finally, it showed how stigmergy and pheromone evaporation can help explore new paths, while also allowing to re-explore previous paths in order to recover from failure-related data losses.

In this paper we study this data-collection problem within complex networks (rather than within uniform spaces – studied previously). This is an important difference, since the topology of the network explored has a significant impact on the agents' performance, as they explore, collect and exchange information [11]. As before, agents may fail at different rates; yet we assume accurate data-collection – i.e. when agents are available their information is reliable (as opposed and complementary to [9]). We study two motion algorithms – random and stigmergy. These are similar to the ones defined in [8]; as Lévy walks do not apply to non-directional spaces, like networks. *The objective is to analyse how agent performance* (i.e. how fast all network data is collected) *and robustness* (i.e. how task completion is achieved in the face of agent failures) *depend on the adopted exploration technique, on the network topology and on failure rates.*

The remaining of this paper is organized as follows. Section 2 presents the data collection problem in complex networks; the studied network topologies; and the agents' design. Section 3 details the analysed motion algorithms, while section 4 presents the experimental settings and discusses obtained results. Conclusions and future works are presented in section 5.

2 The Problem of Data-Collection in Complex Networks

The problem studied can be summarised as follows. Agents must explore a complex network (simulated), in order to collect desired data present in the network vertices. Agents move among interconnected vertices based on a predefined algorithm (section 3), collect data from each visited vertex and exchange their data with any agents that they meet at the same vertex. Additionally, agents can fail over time with probability p_f. The aim is to have at least one agent collect all data from the entire network. The parameters of interest are the *speed* of task-completion and the *success rates* in the presence of agent failures, depending on *network topologies*, *agent motion algorithms* and *failure probabilities*.

The agents' implementation is based on [10]. Each agent is equipped with a set of perceptions $p = \{pheromone, data, current\_node, neighbors, msg\}$; where *pheromone* is a vector in R^n with values in $[0, 1]$, representing the amount of pheromone in the agent's vicinity (i.e. vertices adjacent to the current location) [12]; *data* is the information to collect in the agent's current vertex; *neighbour* returns the *id*s of agents in the same vertex; *msg* stores messages from other agents; and *loc* returns the agent's location (vertex name). Each agent can also perform a set of actions $Actions = \{Move(vertex), Collect, Send(msg), Recv\}$; where $Move(vertex)$ moves the agent to the *vertex* location; *Collect* senses data from the agent's current location and stores it in its local memory; and *Send* and *Recv* enable information exchanges with other agents.

Simulation time is defined via discrete *rounds*. In each round, each agent: senses its local environment (e.g. local data, co-located agents and adjacent vertices); decides on an action (e.g. collect and exchange data, select a neighbouring vertex to move to); and performs the actual action [13, 14]. The simulation ends when at least one agent completes the exploration (i.e. collects all the data) or if all the agents fail. The environment is the complex network to be explored.

In short, a *complex network* consists of a large number of interconnected nodes characterized by non-trivial topological properties – i.e. neither purely regular nor completely random; unlike lattices and random networks. Typical features include relative small distances between nodes, high clustering, or power-law degree distributions (i.e. heavy-tailed) [2]. A more formal definition of complex networks is quite difficult to provide; researchers have focused instead on specific topological metrics and on the kinds of node interconnection rules that produce topologies with distinctive properties [11].

In this paper, a complex network is defined as a graph G with a set of vertices V and a set of edges E: $G = (V, E)$. A probabilistic rule defines the way in which vertices are interconnected when constructing the graph [2, 11]. Hence, complex networks with different topologies can be generated by using different rules of inter-connection. In this paper we evaluate the main types of network topologies identified in the literature, namely, *Small-World*, *Scale-free* and *Community* networks (discussed below). We additionally use more regular topologies for comparison, such as *Forest Hub & Spoke*, *Lattice*, *Line* and *Circle*.

2.1 Small-World Networks

A *Small-World* network is generated by starting from a regular network (in terms of node interconnections) and then rewiring some of these connections in a random way [15]. This type of network features relatively short paths between any network nodes, even in very large networks. In this paper, we use a

Watts-Strogatz model [3,16], with different parameters, to generate Small World networks. We start with a regular ring lattice network with n vertices and k edges per vertex, then rewire each edge with a probability β. The β parameter determines how regular the final network will be: $\beta = 0$ generates a regular network, $\beta = 1$ a random network, and in-between values a Small-World network [15] (Fig. 1).

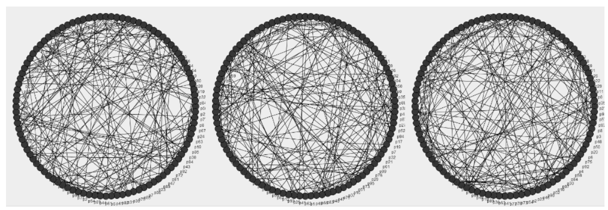


Fig. 1. Small-World Networks: $n = 100$, $k = 4$, a) $\beta = 0.3$, b) $\beta = 0.5$, c) $\beta = 0.9$

2.2 Scale-free Networks

Scale-free networks are characterised by *degree distributions* that follow a mathematical function known as power-law [17]. The *degree distribution* is the probability distribution of the node degrees over the entire network [18]; where a node's *degree* is its number of links. A power-law distribution implies that node degrees may differ by magnitudes of scale, and hence that a few nodes (called *hubs*) have a disproportionate number of links compared to the average degrees. Notable examples of real scale-free networks include the WWW, email or protein interaction networks. They are highly resistant to accidental failures, but rather vulnerable to targeted node attacks [19].

Scale-free networks can be obtained by starting with sn nodes and η connections. At each step a new node is added and connected via η links to existing nodes, based on preferential attachment (i.e. more likely to connect to nodes with higher degrees) [20]. Namely, the probability to connect to an existing node is defined by $p_i = \frac{k_i}{\sum_j (k_j)}$, where k_i is the degree of node i [16,21,22]. The process is repeated for *steps* times [11]. Figure 2 depicts different configurations showing how the number of connections increases with *eta*.

2.3 Community Networks

Community networks feature structures where nodes can be assigned to different groups, or clusters, that are highly interconnected internally, and have relatively few connections among nodes belonging to different groups [23]. In this paper Community networks were generated using a *n_clusters* parameter to define the

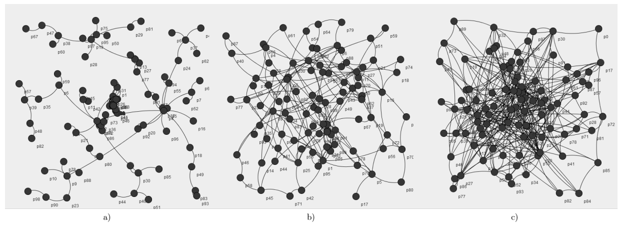


Fig. 2. Scale-free Networks: $sn = 4$, $steps = 97$, a) $\eta = 1$, b) $\eta = 2$, c) $\eta = 4$

number of groups in the network and adding a single connection between nodes of different groups. Each group was generated as a small world network (with its own k, β, and $n = m/n\_clusters$, where m is the number of nodes in the network). Figure 3 shows a Community network with four groups connected either via a central node (selected at random), or via a circle formed by pairs of nodes selected from different groups (also random).

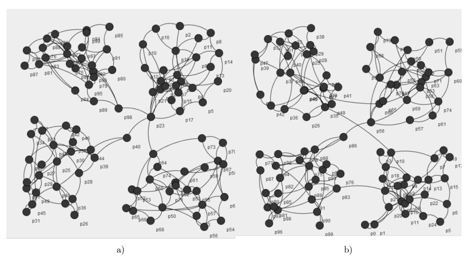


Fig. 3. a) Community Central, and b) Community Circle Network: $n\_clusters = 4$, $\beta = 0.5$, $degree = 4$ and $m = 100$

2.4 Forest Hub & Spoke

The *Forest Hub & Spoke* network is based on the Hub & Spoke (or Star) configuration, where all nodes are connected (spokes) around a central node (hub); the forest is then formed by connecting pairs of such Star structures. This type of network ensures high availability and reliable computing services because it allows expansion of individual cloud instances [24]. In this paper, we generate 4 Hub & Spoke clusters of 25 nodes each, as shown in Figure 4.

2.5 Line, Circle and Lattice

Experimental design includes Lattice, Line and Circle topologies. The purpose of performing experiments with these topologies is to test exploitation and exploration properties of the selected algorithms with a higher diameter (line), a

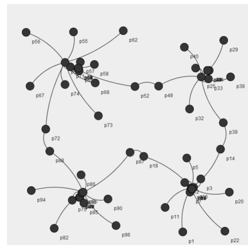


Fig. 4. Forest Hub & Spoke

long path length (line and circle), and regular connections (lattice). Figure 5, shows the configurations applied to the experiments each one with 100 nodes.

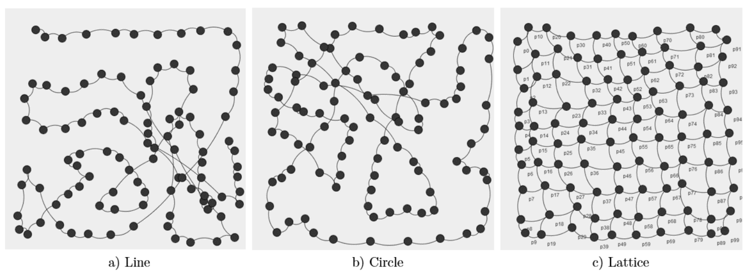


Fig. 5. Line, Circle and Lattice

3 Agent Motion Algorithms and Failures

After selecting different types of complex networks that represent various topologies, we aim to establish and profile how a determined motion strategy influences the exploration of unvisited vertices, the encounter of other agents and robustness given in terms of completing the task even when agents prone to failure; as all of them are important for data collection.

Each agent implements an algorithm which determines how to sense data from the environment and how to select actions such as motions and communication with other agents. The agent program pseudocode is listed in Algorithm 1. Agent failures are also defined in this program, and produced with a failure probability p_f – e.g., $p_f = 0.1$ means the agent fails on average every 1 out of 10 rounds.

Based on perceptions, agents choose the next vertex to move to based on two possible motion processes: random, which is exploratory; and pheromone-based, which improves exploitation of new paths and agent encounters (which in turn enhances collective exploration).

Algorithm 1 Agent program

```
Percept p
Action action
round ← 0
while Agent.status ≠ Fail do
    λ ← U[0,1]                                        ▷ uniform random number
    if λ < p_f then
        Agent.status ← Fail
        break
    end if
    p ← environment.sense()
    Agent.move(motionAlgorithm(p))
    if Agent.hasNeighbors(p.location) then Agent.exchange(p.neighbors)
    end if
    round ← round+1
end while
```

When following a *random walk* an agent selects its moving direction randomly from the set of vertices adjacent to the current vertex, at each round. A uniformly distributed pseudorandom generator is used for generating the random sequence.

When following a *pheromone-based* movement an agent chooses vertices based on their pheromone load (e.g. in order to find unexplored vertices, or other agents). This algorithm is based on the Ant Colony System algorithm (ACS) [25], using stigmergy, and an adaptation of the *Carriers algorithm* in [10]. Initially, all vertices have a pheromone value $\tau_v = 0.5$. As in ACS [25], a random variable $q \in [0,1]$ dictates when to apply an exploitation rule or biased exploration (Eq. 1):

$$dir = \begin{cases} \text{exploitation rule} & \text{if } q \leq 0.9 \\ \text{biased exploration} & \text{otherwise} \end{cases} \tag{1}$$

A carrier agent chooses the direction with the *minimum* pheromone amount in its vicinity, looking for uncharted vertices. If more than one vertex has the same *minimum* value a random direction is picked among these.

Biased exploration is a random-proportional rule [25] which gives an agent i a probability of choosing a vertex $p_d(v)$ depending on the amount of pheromone τ_v in its vicinity $neighbourhood(i)$ (Eq. 2). $neighbourhood(i)$ includes the vertices connected to vertex i. This prevents agents from getting trapped in a confined area (e.g. carriers surrounded by pheromone traces). For carriers $\tau_v'(v) = 1 - \tau_v(v)$.

$$p_d(x,y) = \frac{\tau_v'(v)}{\sum_{(k)\in \text{neighbourhood(i)}} \tau_v(k)} \tag{2}$$

Whenever an agent i moves, at each rount t, it updates its internal pheromone value $\tau_{a_t}(i)$ (as in Eq 3) and also the pheromone amount in its current vertex location $\tau_v(v)$ (as in Eq 4).

$$\tau_{a_t}(i) = (\tau_{a_{t-1}}(i) + 0.01 * (0.5 - \tau_{a_{t-1}}(i))) \quad (3)$$

$$\tau_{v_t}(v) = \tau_{v_{t-1}}(v) + 0.01 * (\tau_{a_{t-1}}(i) - \tau_{v_{t-1}}(v)) \quad (4)$$

If an agent i, finds or receives new information, then its internal pheromone value is updated to $\tau_{a_t}(i) = 1$. In this case, Eq. 3 decreases internal pheromone value at each round; and Eq. 4 increases the amount of pheromone in the locations that the carrier agent explores.

Passive evaporation is added so as to make explored paths less dominant and allow re-exploration of routes of agents that fail without sharing information [26]. This type of evaporation is performed by the environment rather than by agents and it is applied in all the vertices of the complex network G, using the definition in [27] corresponding to Eq. 5, with evaporation rate $\rho = 0.01$:

$$\begin{aligned} \tau_{v_i} = (1-\rho)\tau_{v_{i_{(t-1)}}}, \text{ for} \\ \forall_i \in \{V, G = (V, E)\} \end{aligned} \quad (5)$$

4 Experiments and Results

Experiments aim to analyse the performance of motion algorithms to solve the data-collection problem in complex networks. We apply metrics of speed, amount of information collected versus time, robustness (in terms of failure resistance) and number of messages sent. Additionally, experimental design provides insights regarding impact of a selected complex network in the agents performance and suitability of motion algorithms to achieve the data collection task in a determined topology.

4.1 Experimental settings

Each experiment is defined via a combination of: a different complex network topology (section 2), a different agent motion algorithm (section 3), and a different failure probability p_f for all agents. In all cases, the network consist of 100 vertices and is explored by 10 agents (a relation 10 to 1 from vertex to agent). Each experiment was performed 30 times. Agents start from random locations, selected separately for each topology (but the same ones for all 30 repetitions in any one topology). Each simulation stops when one agent collects all the information from all network vertices, or if all the agents fail. We compare the performance of the two movement algorithms (i.e. Random and Stigmergy-based) within different topologies (listed below), and for different failure probabilities starting from zero and by increasing p_f until a value in which most of the experiments fail (p_f = 0, 0.001, 0.003, 0.005 and 0.008). The specific parameters used for each complex network topology studied are:

– Lattice: size 10×10;
– Small World: degree 4 , $\beta = 0.1$, 0.3, 0.5, 0.9;
– Scale-free: number of steps $steps = 97$, starting nodes $sn = 4$, added links per step $\eta = 1$, 2, 4;
– Community: $\beta = 0.1$, 0.3, 0.5, 0.9, $n\_clusters = 4$;
– Community Circle: $\beta = 0.1$, 0.3, 0.5, 0.9, $n\_clusters = 4$;
– Forest Hub & Spoke: $n\_clusters = 4$;
– Line: no specific parameters;
– Circle: no specific parameters.

4.2 Results and discussion

Agents are evaluated on different criteria in scenarios *with* and *without* agent failure. When agents don't fail ($p_f = 0$) results are analysed terms of the agents' *performance* – i.e. *number of rounds* before the first agent collects all the data. Figure 6 depicts the round numbers for the two algorithms in different network topologies. When agents don't fail, all experiments are ultimately successful. When agents do fail ($p_f > 0$), the agents' *robustness* is evaluated instead in terms of *success rates* – i.e. how often the agents complete the task – and *rate of global data collection* – i.e. how fast the agents collect the data together (rather than individually). Finally, the *number of messages* exchanged among agents is also evaluated, as critical in limited resource environments.

An ANOVA test is also performed for failure-less experiments, to determine whether or not the observed differences between the round number means for the two motion algorithms (Random and Carriers) are statistically significant. The null and alternative hypothesis for a determined topology are the following:

– H_0: round number means for the two algorithms are *equal* for a network G;
– H_1: round number means for the two algorithms are *different* for a network G, indicating a correlation between the algorithm and the round number.

Table 1 shows the ANOVA test results. The *F-value* represents the F statistics – the variation between the round numbers of the two algorithms, in the given network. The *p-val* and *p-wilc* indicate the statistical significance between the results of the two algorithms for all the topologies (since $p\_val < 0.05$ and $p\_wilc < 0.05$) except for the Community network with a $\beta = 0.5$ and $clusters = 4$ (where $p\_val > 0.05$ and $p\_wilc > 0.05$; marked as $*$ in Table 1).

Hence, based on the round number box-plots in Figure 6 and the ANOVA test, we can conclude that Carriers are faster than Random agents for most network topologies, when $p_f = 0$. However, as observed in Figure 6 (and marked as $\star$ in Table 1), Random exploration is faster than the Carriers for the Forest Hub-&-Spoke (Figure 6-e) and the Scale-free with $sn = 4$, $\eta = 1$, $steps = 97$

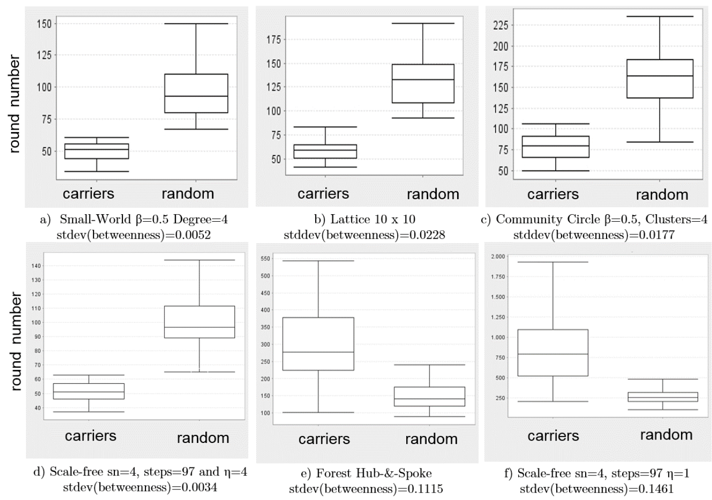


Fig. 6. Box-plot of round number by some selected complex networks with $p_f = 0$

Table 1. ANOVA and Wilcoxon Test for Carriers vs Random, by Topology

Topology	**F-value**	**p-val**	**Dif**	**p-wilc**
Line	111.75	3.806e−15	465	1.824e−6
Circle	248.28	2.2e−16	465	1.822e−6
Lattice 10×10	75.996	2.2e−16	89.7	0.0257
Scale-free				
$(\star) sn = 4$, $\eta = 1$, $steps = 97$	91.807	1.476e−13	2.5	2.352e06
$sn = 4$, $\eta = 2$, $steps = 97$	69.922	1.52e−11	465	1.821e−6
$sn = 4$, $\eta = 4$, $steps = 97$	128.04	2.2e−11	465	1.822e−6
Forest Hub and Spoke				
$(\star)$ $clusters = 4$	19.755	3.573e−5	36.5	1.355e−5
Community Network				
$\beta = 0.1$, $clusters = 4$	47.382	3.952e−9	490	2.21e−6
$\beta = 0.3$, $clusters = 4$	46.56	5.797e−9	444.5	1.359e−5
$(*)$ $\beta = 0.5$, $clusters = 4$	2.304	0.1345	322.5	0.06561
$\beta = 0.9$, $clusters = 4$	18.228	6.84e−5	474	8.928e−5
Community Circle				
$\beta = 0.1$, $clusters = 4$	152.07	2.2e−16	435	2.701e−6
$\beta = 0.3$, $clusters = 4$	144.48	2.2e−16	422.5	9.77e−6
$\beta = 0.5$, $clusters = 4$	93.448	1.07e−13	465	1.822e−16
$\beta = 0.9$, $clusters = 4$	121.477	7.472e−16	465	1.823e−6
Small World				
$\beta = 0.1$, $degree = 4$	126.86	3.149e−16	465	1.821e−6
$\beta = 0.3$, $degree = 4$	65.385	4.394e−11	465	1.822e−6
$\beta = 0.5$, $degree = 4$	85.378	5.34e−13	465	1.823e−6
$\beta = 0.9$, $degree = 4$	144.64	2.2e−16	465	1.817e−6

networks (Figure 6-f). This is probably due to the fact that in these topologies most paths pass through unique large hubs. Therefore, in the Carriers case, these hubs are pheromone-marked very often and hence slow-down agent movement across sub-networks.

We aim to quantify the topological features that impact agent performance. Hence, we tried to identify a correlation between the round number and topological metrics like network diameter, degree distribution, clustering coefficient, and betweenness centrality. We found that the two exception topologies (Scale-free $sn = 4$, $\eta = 1$ and $steps = 97$ and Forest Hub-&-Spoke) feature greater values for the *standard deviation of the node betweenness centrality* value – $stdev(betweenness)$ – compared to other topologies, as in Figure 6 e) and f).

To test this correlation, we generated more Scale-free network instances using the same parameters: $sn = 4$, $\eta = 1$ and $steps = 97$; and plotted their $log(round\ number)$ versus $stdev(betweenness)$. Indeed, Figure 7-a shows the correlation of the betweenness centrality and the round number for all topologies, including the additional Scale-free ones. Since most topologies have relatively low betweenness values (lower than 0.025) compared to the Scale-free cases (greater than 0.05), we only show these cases in Figure 7-b, for clarity, to highlight that a correlation also exists for these topologies, even if at a different scale. Figure 7-c shows the same correlation for the Random algorithm, for all topologies; and Figure 7-d shows the correlation for both algorithms, for all topologies. The Carriers algorithm seems to feature a stronger relation between the round number and the betweenness centrality, compared to the Random case, with greater betweenness values causing larger round numbers (i.e. lower performance). For system designers, this means that selecting the best agent exploration algorithm depends on the network topology (betweenness centrality); and the selected algorithm may have to change over time, for best performance, as the network topology evolves.

We also evaluated the *global information collected by all agents combined* (rather than by each agent). This is useful for analysing algorithm robustness in case of agent failure, especially for applications where all agents can communicate data collected to a central location, and where a percentage of the complete data suffices (e.g. 90%). Hence, the shape of the function describing global information collected in time is important, with steeper shapes offering better robustness, as data is collected faster, before agents start failing. Global information is measured in each experiment (i.e. given topology, motion algorithm and p_f) by reading the local information collected by each agent, at each round, and calculating the total sum. Figure 8 presents the increase of the global information with the round number for the Scale-free network (generated with parameters $sn = 4$, $\eta = 1$ and $steps = 97$) for the two algorithms. Each experiment is performed 30 times and the minimum, median and maximum values plotted. Results show that global information is collected faster by Carriers than Random agents – e.g. at round 50, the minimum collected by Carriers is about 90% whereas by Random is only 70%; at round 100, the minimum for Carriers is 97% and for Random about 85%). It also seems that for Carriers the longest

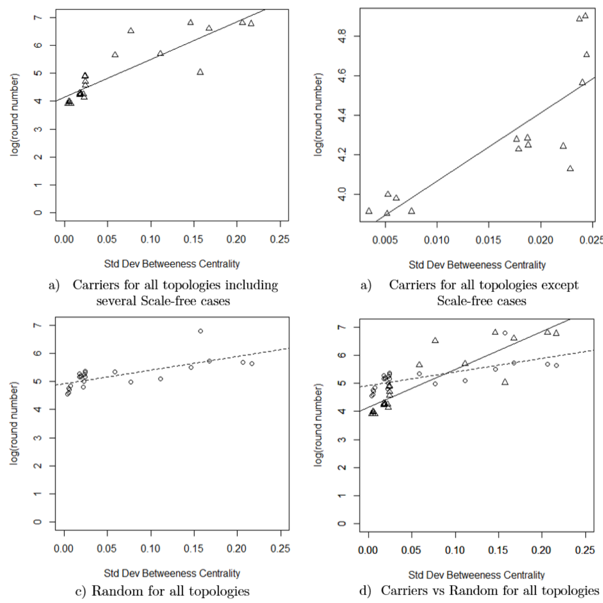
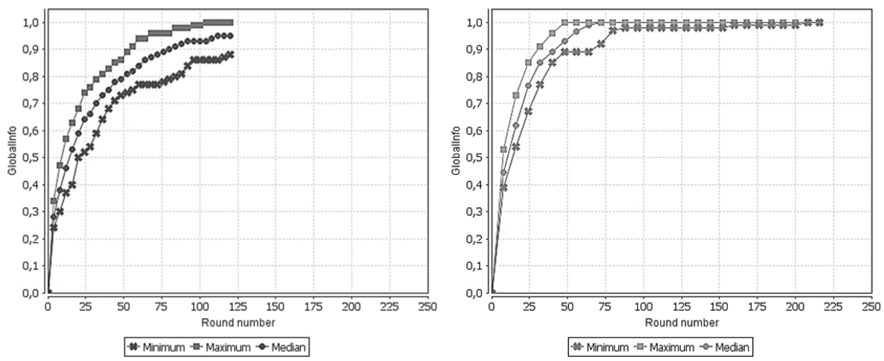


Fig. 7. Correlating the betweenness centrality (std. dev.) to the round number (log)

time is spent for collecting the last 3% of the data, which causes the Carriers to be slower than Random for collecting all data in this topology (Cf. Figure 7-d). This means that in applications where less than 97% of data collection suffices (e.g. some sensor networks) the Carriers can outperform Random agents even for such topologies. Finally, in all cases, the median values approximate the maximum ones much faster for the Carriers than for Random (e.g. round 75 for Carriers and not before round 125 for Random).

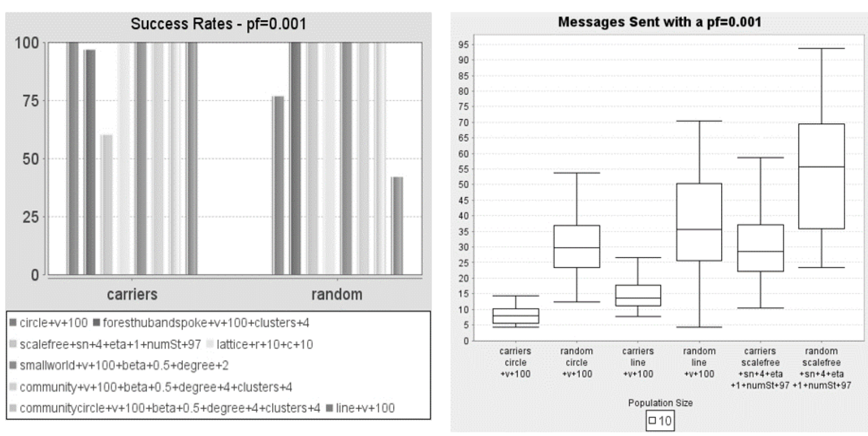
Let us now study the cases where agents can fail. Figure 9 provides a histogram of the success rates (a) and the box-plots of messages sent (b), for selected topologies, and $p_f = 0.001$. Figure 9-a shows that for the Carriers, the topologies most impacted by agent failures are the Scale-free (with $sn = 4$, $\eta = 1$ and $steps = 97$), where success rates drop to about 60%; and the Forest-Hub-and-Spoke, to about 96.66%. For the Random algorithm, the only impacted topologies are Circle and Line. By comparing these success rates with the round number evaluations (Cf. Figure 7), we can note that faster data collection favours success rates. Figure 9-b shows a higher number of message exchanges among Random agents compared to Carriers. This could explain the lower success rates for Carriers in topologies that infringe agent circulation (e.g. some Scale-free cases), since agents are less likely to meet and their information is lost when they fail.

Figure 10 shows the global information collected for the topologies most impacted by agent failures (Scale-free and Circle). In both cases, the median



a) random b) carriers

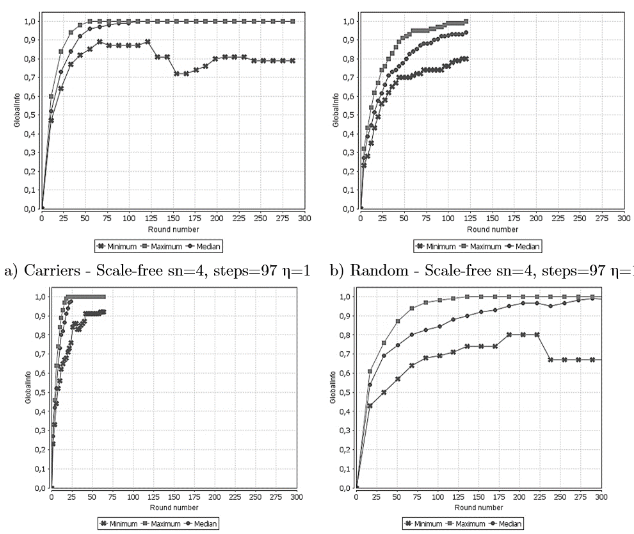
Fig. 8. Global information collected for Scale-free $sn = 4$, $\eta = 1$, $steps = 97$ and $p_f = 0$



a) Success Rates b) Messages Sent

Fig. 9. Success rates and messages sent for $p_f = 0.001$

reaches 100% faster for Carriers than for Random agents. Also, in the Scale free case where success rates suffer, Carrier agents actually manage to collect all the information together, yet they never meet to share the information and hence no single agent completes. Random agents are slow to explore Line and Circle networks as they move around the same vertices and share the same local information. Hence, when they fail, their information is lost and other agents do not reach the same areas before their own failures.

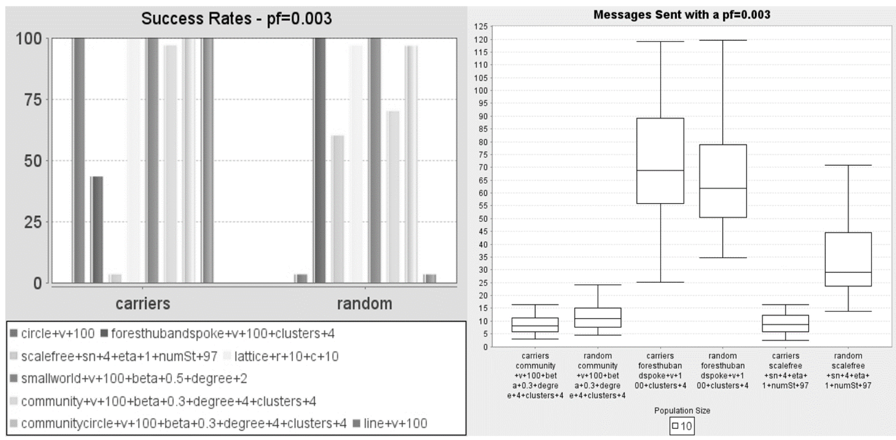


a) Carriers - Scale-free sn=4, steps=97 η=1
b) Random - Scale-free sn=4, steps=97 η=1
c) Carriers - Circle
d) Random - Circle

Fig. 10. Global information collected for Scale-free $sn = 4$ $\eta = 1$ and $steps = 97$ and Circle, with $p_f = 0.001$

Figure 11-a shows that for Carriers with $p_f = 0.003$ the success rate for the Forest-Hub-and-Spoke topology is further reduced, to 43%; and for Scale-free ($sn = 4$, $\eta = 1$ and $steps = 97$) to 4%. Random agents also start featuring lower success for this Scale-free topology (60%). Community networks start suffering in the Random case, while not being impacted when Carriers are used. For Line and Circle networks, Random exploration becomes severely impaired (less than 10%), while Carriers maintain 100% success rates. The success of both algorithms remains intact (100%) for Small-world topologies.

Figure 11-b indicates that for the Community network Random agents exchange more messages than Carriers, even with less success rates, signifying that they probably exchange redundant data. Figures 12-a and 12-b show a fast data collection for Carriers, indicating that Carriers are better for exploitation of new vertices in these networks, since they feature higher success rates despite exchanging fewer messages. In the Forest Hub & Spoke network, agents exchange



a) Success Rates

b) Messages Sent

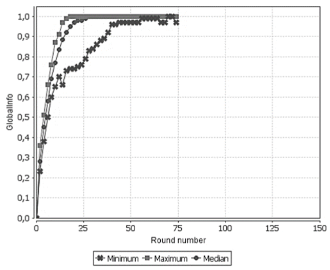
Fig. 11. Success Rates and Messages Sent for $p_f = 0.003$

more messages, via the hubs, and the median of global information converges faster to maximum value for Carriers than for Random.

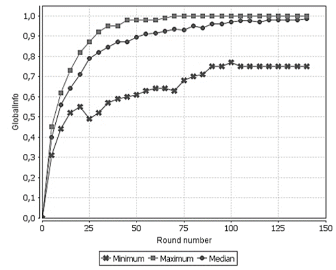
Figure 13 shows success rates for For $p_f = 0.005$ in each kind of network, for different generation parameters. Small World (13-a) and Community Circle (13-c) feature the highest success rates compared to other topologies. Carriers perform well in all Small World networks, whereas Random performs worse (yet relatively better when Small Worlds are generated with higher beta values – less regular and more random graphs) – Figure 13-a. Carriers also start to decrease success rates in Community networks (Figures 13-b and 13-c), especially when clusters are less regular (i.e. greater Beta values). Also, success in Community networks is lower than in Community Circle, for both algorithms, since the clusters are connected through a single central vertex, impeding movement.

For a $p_f = 0.008$, only the Carriers manage to reach success rates of over 70%, for the most failure-resistant topologies: Lattice, Small World (all configurations), Community Circle (all configurations) and Scale Free (4-4-97).

Additionally, a statistic test was performed for the Community network ($\beta = 0.5$) and a $p_f = 0.001$, because there was not a statistically significant difference in terms of round number for $p_f = 0$, while the box-plot for the round number does show a difference (Figure 14-a). By taking advantage of the 100% success rates for both algorithms, a Wilcoxon test for round number and a $pf = 0.001$ indicated a $p\_val = 0.0001538$. This means that there is a significant difference between the round number means, with Carriers being faster than Random, when failures occur. For the other failure probabilities success rates decrease faster for Random agents than for Carriers (Figure 14 b, c and d).

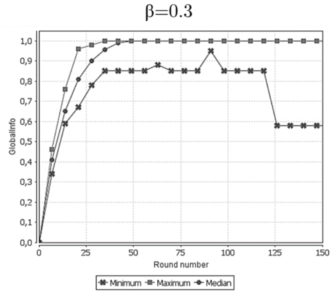


a) carriers – Community Network

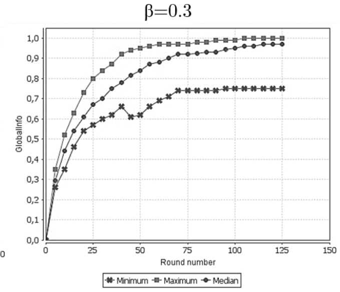


b) random – Community Network

β=0.3 β=0.3

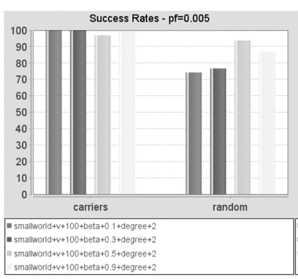


c) carriers – Forest Hub-&-Spoke

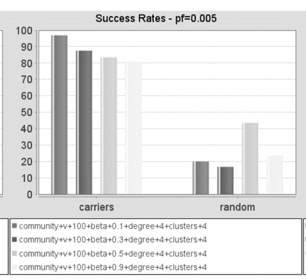


d) random – Forest Hub-&-Spoke

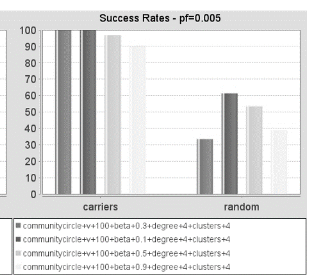
Fig. 12. Global Information Collected for Community Network $\beta = 0.3$ and Forest Hub-and-Spoke, with $p_f = 0.003$



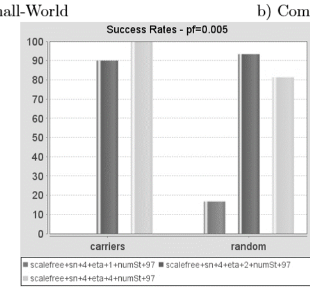
a) Small-World



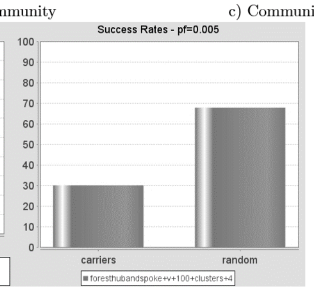
b) Community



c) Community-Circle



d) Scale-free



e) Forest Hub-&-Spoke

Fig. 13. Success Rates For Complex Networks $p_f = 0.005$

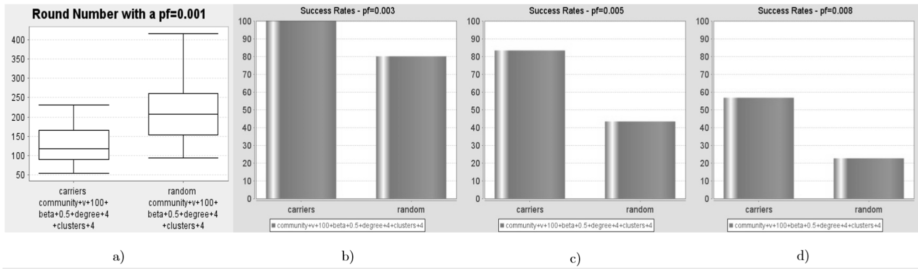


Fig. 14. Community Network Beta=0.5. a) Box-plot of round number with $p_f = 0.001$, b), c) and d) Success rates for $p_f = 0.003, 0.005$ and 0.008 respectively

5 Conclusions and Future Work

In this paper we studied the problem of data-collection in complex networks using failure-prone agents. We evaluated two agent motion algorithms (random and a pheromone-based) for exploring complex networks with different topologies: Small-World, Scale-Free and Community Networks. We also studied several regular network topologies for comparison purposes: Lattice, Forest Hub & Spoke, Line and Circle. Experimental results showed that a pheromone-based exploration technique improves the exploitation of new paths and results in faster data-collection for most experiments in the different topologies. Results also indicate a relation between network topology and data-collection performance, where the differentiating factor among topologies can be quantified as the variance of the *betweenness centrality* among nodes. Namely, the higher the standard deviation of the betweenness centrality of nodes in a complex network, the higher the completion times of the data-collection task in that network.

As shown in [10], the agents' success in completing their collective task relies critically on each agent's motion process. This result was also confirmed here for the case where the space was a complex network (rather than a uniform surface). Also, a faster exploration algorithm provides better resistance to agent failure.

The Carriers strategy is also good for exploitation of new vertices and for reducing the number of messages exchanged. Indeed, the median of the global information collected (by all agents together) was approximated to 100% faster than in the random case, in all the experiments. In the experiments where the random approach was more successful (i.e. Scale-free topology generated with $sn = 4$, $\eta = 1$ and $steps = 97$ parameters; and Forest Hub & Spoke topology) it was difficult for the carrier agents to meet and exchange their local information (as pheromone-marked hubs prevented circulation among sub-networks they interconnected). However, on a global level, carrier agents collect information faster than random agents in all scenarios (even with failures). Small-worlds, Community Circle Networks, Lattice and Scale-free with a degree of 4 (4-4-97) are faster than the other topologies for data collection, for both algorithms (random and carriers), with the Small-world being the fastest topology.

Future work will study various network scales and agent population sizes, and introduce node failures. The case where agents aim to synchronise different data versions among network nodes will also be studied.We believe that obtained results provide key information on the characteristics of different decentralised data-collection algorithms, depending on their application context (e.g. network topology and failure rates). This, in turn, allows system designers to select the best option for their particular application and execution environment, covering a broad spectrum of applications like sensor networks, swarm robotics, server clusters, clouds, systems of systems and the Internet of Things (IoT).

References

1. Tanenbaum, A., Steen, M.V.: Distributed systems: principles and paradigms. Prentice-Hall (2006)
2. Boccaletti, S.: The Synchronized Dynamics of Complex Systems. Elsevier, Florence, Italy (2008)
3. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature **393** (1998) 440–2
4. Grabow, C., Hill, S.M., Grosskinsky, S., Timme, M.: Do small worlds synchronize fastest? EPL (Europhysics Letters) **90** (2010) 48002
5. Noh, J.D., Rieger, H.: Random Walks on Complex Networks. Physical Review Letters **92** (2004) 118701
6. Ongaro, D., Ousterhout, J.: In Search of an Understandable Consensus Algorithm. Ramcloud.Stanford.Edu (2013)
7. Nedic, A., Ozdaglar, A., Parrilo, P.a.: Constrained consensus and optimization in multi-agent networks. IEEE Transactions on Automatic Control **55** (2010) 922–938
8. Rodriguez, A., Gomez, J., Diaconescu, A.: Towards Failure-Resistant Mobile Distributed Systems Inspired by Swarm Intelligence and Trophallaxis. In: Proceedings of the European Conference on Artificial Life 2015, At The University of York UK. (2015) 448–455
9. Nguyen Vu, Q.a., Hassas, S., Armetta, F., Gaudou, B., Canal, R.: Combining trust and self-organization for robust maintaining of information coherence in disturbed MAS. Proceedings - 2011 5th IEEE International Conference on Self-Adaptive and Self-Organizing Systems, SASO 2011 (2011) 178–187
10. Rodriguez, A., Gomez, J., Diaconescu, A.: Foraging-Inspired Self-Organisation for Terrain Exploration with Failure-Prone Agents. In: 2015 IEEE 9th International Conference on Self-Adaptive and Self-Organizing Systems. Volume 2015-Octob., IEEE (2015) 121–130
11. Van Der Hofstad, R.: Random Graphs and Complex Networks Vol. I. Available on http://www. win. tue. nl/rhofstad/NotesRGCN. pdf **I** (2016)
12. Gray, L.: A Mathematician Looks at Wolfram ' s New Kind of Science. Notices of the AMS **50** (2002) 200–211
13. Balaji, P.G., Srinivasan, D.: An introduction to multi-agent systems. Studies in Computational Intelligence **310** (2010) 1–27
14. Russell, S., Norvig, P.: Inteligencia Artificial. Un enfoque moderno. 2da Edición. (2004)
15. Mori, H., Uehara, M., Matsumoto, K.: Parallel Architectures with Small World Network Model. 2015 IEEE 29th International Conference on Advanced Information Networking and Applications Workshops (2015) 467–472

16. White, S.: Analysis and visualization of network data using JUNG. Journal Of Statistical Software **VV** (2005) 1–35
17. Barabási, A.: Emergence of Scaling in Random Networks. Science **286** (1999) 509–512
18. Gündüz-Öüdücü, u., Etaner-Uyar, A.i., eds. Lecture Notes in Social Networks. In: Social Networks: Analysis and Case Studies. Springer Vienna, Vienna (2014)
19. Barabási, A.L., Bonabeau, E.: Scale-Free Networks. Scientific American **288** (2003) 60–69
20. Li, L., Alderson, D., Doyle, J.C., Willinger, W.: Towards a Theory of Scale-Free Graphs: Definition, Properties, and Implications. Internet Math **2** (2006) 431–523
21. Small, M.: "Scale-Free Network" - MathWorld-A Wolfram Web Resource (2016)
22. Takemoto, K., Oosawa, C. In: Introduction to Complex Networks: Measures, Statistical Properties, and Models. John Wiley & Sons, Inc., Hoboken, NJ, USA (2012) 45–75
23. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. Proceedings of the National Academy of Sciences of the United States of America **99** (2002) 7821–6
24. Mahmood, Z., Hill, R., eds.: Cloud Computing for Enterprise Architectures. Computer Communications and Networks. Springer London, London (2011)
25. Dorigo, M., Gambardella, L.M.: Ant Colony System: A cooperative learning approach to the traveling salesman problem. IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION (1997)
26. Bell, J.E., McMullen, P.R.: Ant colony optimization techniques for the vehicle routing problem. Advanced Engineering Informatics **18** (2004) 41–48
27. Dorigo, M., Stutzle, T.: Ant colony optimization. Volume 1. MIT Press (2004)