The Relationship between Emergence of the Shortest Path and Information Value using Ant-like Agents

Colin Chibaya
Department of Computer Science
Rhodes University
Grahamstown, South Africa
c.chibaya@ru.ac.za

Shaun Bangay
Department of Computer Science
Rhodes University
Grahamstown, South Africa
s.bangay@ru.ac.za

ABSTRACT

Ant-like agents forage between two points. These agents' probabilistic movements are based on the use of two pheromones; one marking trails towards the goal and another marking trails back to the starting point. Path selection decisions are influenced by the relative levels of attractive and repulsive pheromone in each agent's local environment. Our work in [5] evaluates three pheromone perception strategies, investigating path formation speed, quality, directionality, robustness and adaptability under different parameter settings(degree of randomness, pheromone evaporation rate and pheromone diffusion rate). We re-evaluate two of these strategies in terms of the amount of information they provide using Shannon's formulation [3, 4, 8, 9, 12, 14, 15, 16, 17]. We determine information as the difference between uncertainty before and after path selection decisions. Our focus in this paper is on investigating relationships between the emergence of the shortest path and the amount of stigmergic information that exists in the form of pheromone. Agents are deployed centrally and emergence measures are determined using the worst, reference and best cases observed in [5]. Additionally, the amount of local and global information that is available to agents in each movement step is evaluated. Furthermore, Pearson's correlation coefficients between measures of emergence and the amount of information are calculated. The significance of these correlation coefficients is tested using a 2 tailed test at 1% level of significance. Consequently the relationship between the amount of information and emergent behaviour is established. Significant relationships between information and the emergence of the shortest path exist when strong emergent behaviour is present.

Categories and Subject Descriptors

H.3.3 [**Information retrieval**]: Information theory, search, quantification, Algorithms and interaction techniques

Keywords

Information , entropy, uncertainty, emergence, shortest path, correlation, significance level, similarity, distance.

1. INTRODUCTION

Biologically inspired swarm models such as cellular automata and ant colony optimization are used in many areas of search and optimization [1]. The emergent behaviour resulting from these computational models provides benefits in terms of efficiency, quality of services [6], robustness, adaptability and scalability. However there are as yet no methods for quantifying the level of emergent behaviour and consequent level of performance of a given model. This is unsatisfactory not only from a methodological but also from a pragmatic point of view: if the methods for quantifying the level of emergent behaviour are not investigated, then some under-performing models are likely to go unnoticed. We hypothesize that the amount of *information* that is made available to agents provides a measure of the degree of emergence to be accomplished, and investigate the validity of the claim with regard to the path formation problem using ant algorithms.

Our work in [5] investigates and evaluates three pheromone perception strategies for controlling ant-like virtual agents foraging between two points; a nest-like starting point and a food-like goal. All agents are identical, simple, autonomous and can only communicate locally and indirectly through the environment. This communication mechanism is known as stigmergic communication. Movement of agents is based on the use of two pheromones; one marking trails towards the goal and another marking trails back to the starting point. In every agent's movement step, there are eight path selection possibilities. One of these is chosen probabilistically based on pheromone concentrations in each of the neighbouring cells. Conceptually, uncertainty (as an information theoretic concept) occurs when there are several path selection possibilities within an agent's local environment and we do not known which one will be selected [8]. Ideally, neighbouring locations with the best attractive and repulsive pheromone balances would be selected. However, there are no rules to avoid unexpected (surprise) selections. Such an environmental mechanism for coordinating agents' movements based on probabilistic path selection decisions is called information foraging [4].

We re-assess the performance of successful pheromone perception strategies (A and B in [5]) in terms of the amount

of information they provide. After quantifying measures of emergence (speed, quality, directionality, robustness and adaptability) using the worst, reference and best cases observed in [5], we compare these to the amount of both local and global information that is available to agents.

1.1 Problem statement

We investigate the following measures of emergence (defined in sections 3.2 and 3.3):

Speed/Quality: Path formation speed measures the influence of both attractive and repulsive pheromone. Path quality on the other hand indicates the tendency of agents to follow the path as opposed to reaching the goal by random guess [5].

Directionality: This metric measures the frequency with which agents correctly orient towards the desired target when they intersect with the emergent shortest path [5].

Robustness/Adaptability: Robustness evaluates path formation speed and path quality in environments that consist of obstacles. On the other hand, adaptability refers to a situation where an even shorter path is re-established after one or all obstacles are removed from the environment where a path has already formed [5].

Measures of the amount of local and global information that is available to agents during path selection decisions are equally important. We quantify:

Local information: This is the difference between uncertainty before and after path selection decisions with respect to an agent's local environment.

Global information: Is the gap between uncertainty before and after path selection decisions in respect to the entire terrain.

Our goal is to determine the degree of correlation between the measures of emergent behaviour (path formation in this instance) and measures of information.

1.2 Overview

The rest of the paper is structured as follows: In section 2, we present related work concentrating on different techniques that have been suggested for quantifying information. Path formation literature is intensively covered in [5] and is not repeated here. Our emphasis is on information quantification techniques that are applicable in the evaluation of goal searching models. We describe the experimental process in section 3, providing details on the measures used and explaining the methods used for testing for correlation. The results are presented and discussed in section 4. We conclude this paper in section 5, highlighting contributions and future directions of our work.

2. RELATED WORK

Emergent properties are key features in most agent based models. The success or failure of a model is normally rated in terms of the quality of emergent behaviour it achieves. It is therefore critical that we can evaluate agent based models in terms of the quality of emergent behaviour. We believe that the amount of information present serves as a predictor of emergent behaviour.

Traditionally measurement of information in models has been centred on decomposing the model into its component parts [3, 10, 14, 15]. In this approach models with more component parts are rated as providing more information. But what really counts as distinct component parts in emergent systems? It is known that some properties of emergent systems cannot be represented at component level because the whole is not the sum of its parts.

On the other hand Kolmogorov and Chaitin's algorithmic information theory [2, 3, 4, 8, 9, 12, 15] is also popular. This theory, commonly known as the Kolmogorov complexity (or algorithmic entropy or program-size complexity), measures information using the amount of computational resources that are required to generate the model. The length of the computer program that would generate the model indicates the amount of information that the model achieves. Unfortunately this measure is not easily applied in the context of emergent systems.

Closely related to Kolmogorov's information theory is the logical depth complexity [3, 15], where information is predicted by the run time needed to execute the algorithm that defines the model. The longer it takes to run a model, the more information that model has associated with it. Similarly, information may be defined by the number of arguments declared in the algorithm [10]. This is analogous to Feldman and Clutchfield's theory [15] where information is directly related to the amount of memory stored in the system's configuration. Regrettably, these measures are generally uncomputable [3] and subjective.

Our work is based on Shannon's formulation [3, 4, 8, 9, 12, 14, 15, 16, 17]. In this theory, information is defined as the difference between subsequent entropy measures, one before and another after a path selection decision. As an alternative, changes in degrees of unexpectedness (surprise) in agents' path selection decisions can also provide a measure of information. In this paper we concentrate on information quantification using Shannon's entropy. This quantity is easy to calculate using the stigmergic data that is available in the environment in form of attractive and repulsive pheromone levels.

3. METHOD

We perform simulations of ant algorithms designed to achieve path formation. We collect measurements of both emergence properties in terms of the performance of the path formation process, and information present. The correlation between these two sets of measurements is used to evaluate the extent to which the available information corresponds to the level of emergent behaviour achieved.

3.1 Simulation environment

Our main object in this model is the terrain, a two dimensional square grid of discrete cells. We use the symbol Ω to denote the entire terrain and κ_i to denote a single grid cell, or location. The cardinality of Ω, denoted as $\|\Omega\|$, indicates the total number of locations κ_i in the entire terrain.

Another important object of our model are the myriad ant-like agents that are initially deployed from a single starting point. An agent that is currently at location κ_i is denoted as $a_j(\kappa_i)$. Let $\kappa_n \in \Omega$ be a set of all locations that are within $a_j(\kappa_i)$'s local neighbourhood. We call κ_n the set

of possibilities or scheme of choices for agent $a_j(\kappa_i)$. The cardinality of κ_n is a value less than or equal to 8, thus $\|\kappa_n\| \leq 8$. There are only two cases when $\|\kappa_n\|$ is strictly less than 8; when some elements of κ_n fall off the terrain Ω or when obstacles are included within κ_n.

Upon arrival at a new location say κ_x within the neighbourhood κ_n, an agent $a_j(\kappa_x)$ deposits one of the two possible types of pheromone (search or return pheromone) depending on its internal state. Searching agents release return pheromone which marks trails that other agents may follow towards the starting point. On the other hand, returning agents would release search pheromone, marking trails that provide directional clues towards the goal. The type of pheromone that an agent is currently releasing has repulsive properties to that agent, while the other type has attractive properties. The parameters η and τ respectively denote the amount of repulsive and attractive pheromone levels co-existing in each location. Agents select one destination location within κ_n randomly [5], based on $P_{a_j(\kappa_x)}$. The probability that agent $a_j(\kappa_i)$ would move to a particular $\kappa_x \in \kappa_n$ is defined as:

- Strategy A:

$$Q_{a_j(\kappa_x)} \quad \leftarrow \quad \left(\frac{\tau_x}{\sum_{y \in \kappa_n} \tau_y}\right) - \left(\frac{\eta_x}{\sum_{y \in \kappa_n} \eta_y}\right) \quad (1)$$

$$P_{a_j(\kappa_x)} \quad \leftarrow \quad \lambda \left(Q_{a_j(\kappa_x)} - P_{a_j(\kappa_n),min}\right) + \frac{1-\lambda}{\|\kappa_n\|}(2)$$

- Strategy B:

$$P_{a_j(\kappa_x)} \leftarrow \lambda \left(\frac{\left(\frac{\tau_x}{\sum_{y \in \kappa_n} \tau_y}\right)}{\left(\frac{\eta_x}{\sum_{y \in \kappa_n} \eta_y}\right)}\right) + \frac{1-\lambda}{\|\kappa_n\|} \quad (3)$$

In each of these strategies, the desirability $P_{a_j(\kappa_x)}$, of each κ_n is such that agents favour high quantities of attractive pheromone while penalizing high quantities of repulsive pheromone. In strategy A, the desirability $P_{a_j(\kappa_x)}$ is measured as the difference between levels of attractive and repulsive pheromone relative to the smallest level. On the other hand, the desirability of each κ_n when we use strategy B is measured as the quotient of levels of attractive and repulsive pheromone. These strategies are both meant to favour movements towards high attractive and low repulsive pheromone concentrations. In all cases, pheromone levels are initialized to a very small $\delta = 0.00001$ to avoid any divisions by zero. The parameter $\lambda \in [0, 1]$ denotes a degree of randomness to be associated with each agent's movement decisions irrespective of the level of pheromone present.

3.2 Emergence Measures

Measures of emergence are based on the path formation measures defined in [5].

Speed/Quality is the frequency with which agents arrive at their goals. The time gaps between successive arrivals indicate how fast the emergent path forms as well as the tendency of agents to follow that path as opposed to reaching the goal by random guess. We measure speed by recording the number of successful trips of agents towards their targets in a set time period, in this case 2000 iterations. Shorter time intervals between arrivals correspond to higher speed and use of better quality path. Speed is calculated from the reciprocal values of these time gaps.

On the other hand directionality is measured by first presetting a pheromone path along the line segment that joins the goal and the starting point. Search pheromone along this path is increased linearly towards the starting point and return pheromone increases towards the goal. Two thousand agents are recruited one after the other and the frequency with which each agent correctly orients towards the desired target is recorded. In this work, we use 1 to indicate successful orientation and 0 to indicate failure. Precisely, failure covers the case of wrong orientation or agents crossing over or retreating from the emergent path. Our recruited agents are allowed to modify the preset path by adding more pheromone both on recruitment and in steps that follow thereafter.

Robustness/Adaptability is a variation on path formation speed assessed in environments that contain obstacles. We progressively increase terrain complexity by introducing obstacles one after the other after path formation has already been achieved. The speed with which a new path is formed around these obstacle as well as the quality of that path are recorded, indicating the model's robustness. These obstacles are later removed and the rate with which an even shorter path emerges is measured, indicating model adaptability.

3.3 Information Measures

When evaluating path formation and information measures, agents' local as well as global perceptions are considered. The letters H_l and H_g (adopted from Shannon's notation) respectively denote local and global agent entropy. If H_{l_a} and H_{l_b} denote a particular agent $a_j(\kappa_i)$'s local entropy measures before and after a path selection decision, the absolute difference between the two entropy would indicate the amount of local information available to that agent. As a result, the notation $I_l = H_{l_b} - H_{l_a}$ and $I_g = H_{g_b} - H_{g_a}$ would denote local and global information respectively.

Uncertainty arises because we do not know which one of the locations in κ_n is going to be selected when the agent makes its next path selection decision. In this work, uncertainty and Shannon's entropy measure are synonymous. We determine uncertainty by the number of "clever" questions [8] one can ask before correctly picking the agent's next destination. We start by dividing the solution space κ_n into two equal or almost equal subsets, κ_{n1} and κ_{n2} such that $\|\kappa_{n1}\| \approx \|\kappa_{n2}\| \approx \frac{\|\kappa_n\|}{2}$. A "clever" question would ask if the agent's choice falls in κ_{n1} or κ_{n2}. This approach is similar to the binary searching technique and would minimize the expected, or average number of questions required to make a correct guess by more than half. If we progressively divide the subsets that are formed until their cardinality are one, the process will eventually end after y steps. The number y is nothing else than the logarithm of the cardinality of the set of possibilities in base 2 [12, 16]. Thus, if the set of possibilities is represented as κ_n, then $y = log_2\|\kappa_n\|$.

Furthermore, locations in our solution spaces have associated probabilities of being selected depending on the levels of attractive and repulsive pheromone they contain. Each solution space, say κ_n, together with the set of corresponding probabilities ρ_n, form what is called a probabilistic choice system. It is known from coding theory that the expected uncertainty measures we get in such probabilistic choice systems would reduce to Shannon's entropy measure [19], denoted as $H = -\Sigma_{i=1,n}\rho_i log_2 \rho_i$ [8, 12, 16].

With this, we are able to determine the entropy associated

Local Uncertainty	Global Uncertainty
$P_{a_j(\kappa_n)}$ $\quad \lambda\left(Q_{a_j(\kappa_x)} - P_{a_j(\kappa_n),min}\right) + \frac{1-\lambda}{\|\kappa_n\|}$	$P_{a_j(\kappa_n)}$ $\quad \lambda\left(Q_{a_j(\kappa_x)} - P_{a_j(\kappa_n),min}\right) + \frac{1-\lambda}{\|\kappa_n\|}$
$Q_{a_j(\kappa_x)} \leftarrow \left(\frac{\tau_x}{\sum_{y\in\kappa_n}\tau_y}\right) - \left(\frac{\eta_x}{\sum_{y\in\kappa_n}\eta_y}\right)$	$Q_{a_j(\kappa_x)} \leftarrow \left(\frac{\tau_x}{\sum_{y\in\kappa_n}\tau_y}\right) - \left(\frac{\eta_x}{\sum_{y\in\kappa_n}\eta_y}\right)$
$\lambda\left(\left(\frac{0.0001}{0.0008}\right) - \left(\frac{0.0001}{0.0008}\right)\right) + \frac{1-\lambda}{\|8\|}$	$\lambda\left(\left(\frac{0.0001}{0.1}\right) - \left(\frac{0.0001}{0.1}\right)\right) + \frac{1-\lambda}{\|10000\|}$
$\lambda(0) + \frac{1}{8}$	$\leftarrow \lambda(0) + \frac{1-\lambda}{\|10000\|}$
$\therefore H[P] \quad = -\Sigma_{i=1,m}\rho_i log_2\rho_i$	$\therefore H[P] \quad = -\Sigma_{i=1,n}\rho_i log_2\rho_i$
$= -\sum_{i=1,8}\left(\frac{1}{8}log_2\frac{1}{8}\right)$	$= -\sum_{1,10000}\left(\frac{1}{10000}log_2\frac{1}{10000}\right)$
$= -log_2 2^{-3}\ bits$	$= -log_2 2^{-13.2877}\ bits$
$= 3.000$ bits of information	$= 13.2877$ bits of information

Table 1: **Local and global uncertainty in the first agent 's path selection decision**

Strategy			Worst case			Reference case			Best case		
			λ	evaporation	diffusion	λ	evaporation	diffusion	λ	evaporation	diffusion
	A		50%	0.05%	1%	0%	0%	0%	0%	0%	0.05%
	B		50%	0.05%	1%	0%	0%	0%	50%	1%	0.05%

Table 2: **parameter setup**

with each path selection decision both locally and globally. For example, when the first agent leaves the starting point, there has not been any pheromone deposits in the terrain yet. The level of uncertainty associated with this agent's path selection decisions in this probabilistic choice system are calculated as shown in table 1. In these examples, the amount of both attractive and repulsive pheromone is still 0.0001 (initial quantities that are on terrain before any agent is deployed). The local desirability of each one of the eight neighbouring locations would be $\frac{1}{8}$, which is the probability we use to calculate local entropy. On the other hand, global entropy is calculated as the desirability of a particular location in respect to the entire terrain, thus 10000 (because our grid is 100×100), which gives $\rho_i = \frac{1}{10000}$ before any pheromone has been deposited onto the terrain.

3.4 Correlation

Path formation and information measures are recorded and correlation statistics between each pair of data sets are computed. Commonly, Pearson's Product-Moment correlation coefficient (r) is used, especially when the aim is to measure the strength and direction of linear relationships between the data sets.

Scatterplot diagrams allow visual assessment, particularly of the effects of outlier values. In these, a wide scatter indicates lack of a relationship. On the contrary narrow or elliptic scatterplots indicate stronger correlation. We use scatterplots in section 4.3.1 to verify the relationships we achieve.

Pearson's correlation coefficients between each path formation metric and specific information measures are computed. The levels of significance of each correlation coefficient is tested at a 1% level of significance, using a two-tailed test with $n-2$ degrees of freedom.

Let the variables SQ, D, and RA denote path formation speed/quality, directionality and robustness/adaptability respectively. Correlation coefficients (r) between each path formation measure (SQ, D, and RA) and information measures (I_l and I_g) are calculated using the formula:

$$r = \frac{n\sum xy - \sum x\sum y}{\sqrt{\left(n\sum x^2 - \left(\sum x\right)^2\right)\left(n\sum y^2 - \left(\sum y\right)^2\right)}} \quad (4)$$

where x is either I_l or I_g and y is SQ, D or RA [11]. The significance tests we conduct indicate whether these correlation coefficients are real and not by chance. The value of r ranges from -1, a perfect negative correlation to $+1$ which indicates a perfect positive correlation. A value of r greater than 0.7 is considered to be a strong positive correlation. This is because squares of r indicate the percent of variation in one variable that is related to the variation in the other [18, 7] . Thus, $r > 0.7$ means more than 49% of the variation is related (0.7 squared $= 0.49$).

A critical correlation value read from statistical tables [13] is used to test the significance of the correlation coefficient. For our data ($n > 100$, at a 1% level of significance), this critical correlation value is 0.254. If the absolute value of correlation coefficient we get is greater than the critical correlation value, we conclude that our results are significant and the probability is small that the relationship happened by chance.

3.5 Experimental scenarios

We conduct simulation experiments that produce emergence and information measures for pheromone perception strategies A and B. Each strategy can potentially be parametrized by a large number of parameters, such as those explored in [5]. To avoid the additional complexity of adding these variables, we present results corresponding to the worst,
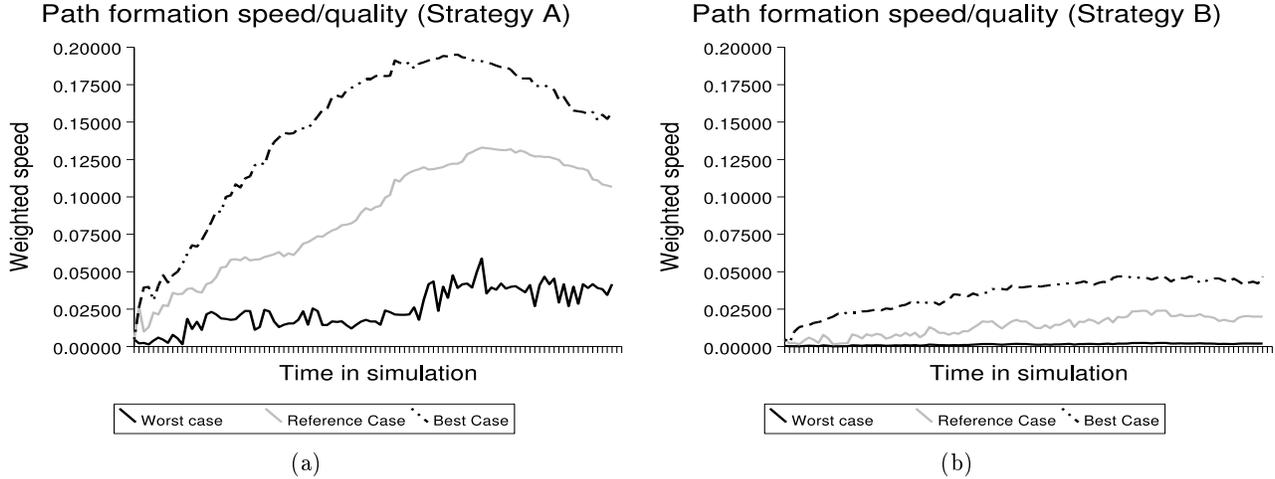
Figure 1: Path formation speed /quality

reference and best parameter settings discovered during this previous work. The parameter settings for these cases are indicated in Table 2.

4. RESULTS

4.1 Emergence Measures

Figure 1 presents the path formation speed measured for strategies A and B when using the worst, reference and best case parameters.These results are consistent results with those achieved in [5] in that agents using strategy A perform best in all categories we measure. Some degree of randomness persists in agents that use strategy B, resulting in loss of path following behaviour. Strategy A performs best when pheromone evaporation is disabled and diffusion is set to 0.05% per movement step. Strategy B perform best when evaporation rate is high because that eliminates stray candidate paths that are created as a result of the randomness in agents' path selection decisions. Path formation speed initially increases with time. However, that path depletes late in simulation time due to pheromone dissipation and terrain saturation.

Figure 2 depicts the frequency with which agents correctly orient towards the desired goal as measures of path directionality. The difference between directionality measures in strategies A and B is initially small because the concentration of attractive and repulsive pheromone at the point where agents are recruited is preset to equal quantities. After a while, the quality of the path modified by agents that use strategy B drops, indicating path depletion due to pheromone dissipation and the degree of randomness that is persistent in these agents' path selection decisions. The quality of the path modified by agents that use strategy A improves, indicating emergence of path following behaviour.

The robustness of the two strategies shows similar behaviour. The speed with which a new path is formed around these detouring obstacles is of our interest. Agents that use strategy A successfully negotiate a new path around obstacles. However, the path that had partially formed when agents use strategy B becomes completely depleted before
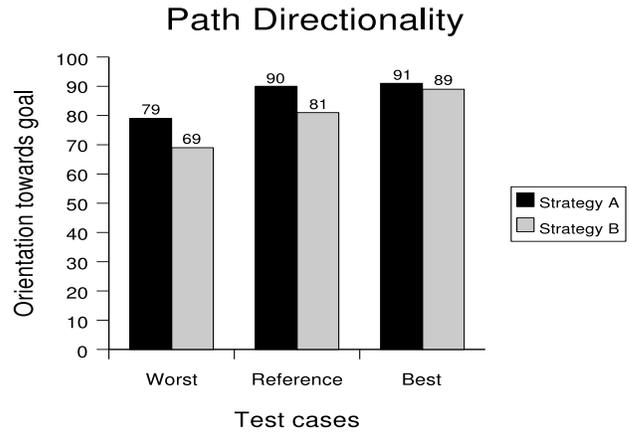

Figure 2: Path directionality

a new path emerged around the obstacle. In other words, obstacles cause agents that use strategy B to revert back to random wandering.

Removal of these obstacles after a new path has formed yields further interesting results. Agents that use strategy A form an even shorter path through those locations where obstacles were, indicating self repair and model adaptability. Unfortunately, agents that use strategy B could not re-establish a path after the inclusion of obstacles. Patches of pheromone deposits form haphazardly around the the terrain revealing the weaknesses of this strategy.

These results show that various levels of path formation formation occur. Thus the conditions simulated correspond to different levels of emergence.

4.2 Information Measures

Uncertainty or entropy measures are high early in simulation time because there is low or no information at all in the terrain. Nevertheless, as more and more agents discover the goal, the amount of both local and global information
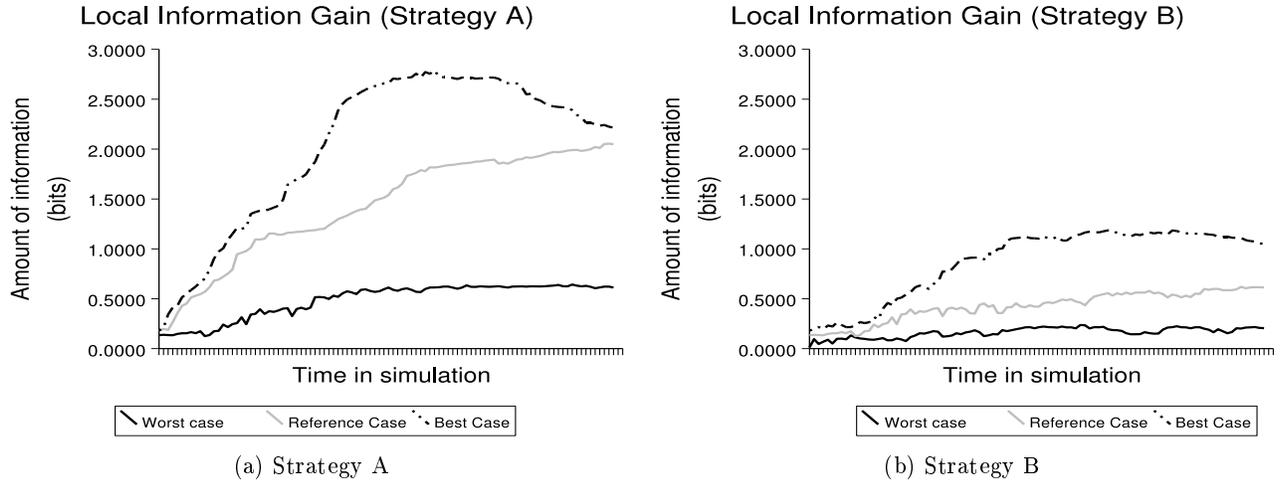
(a) Strategy A   (b) Strategy B

Figure 3: Amount of local information gain

(a)   (b)
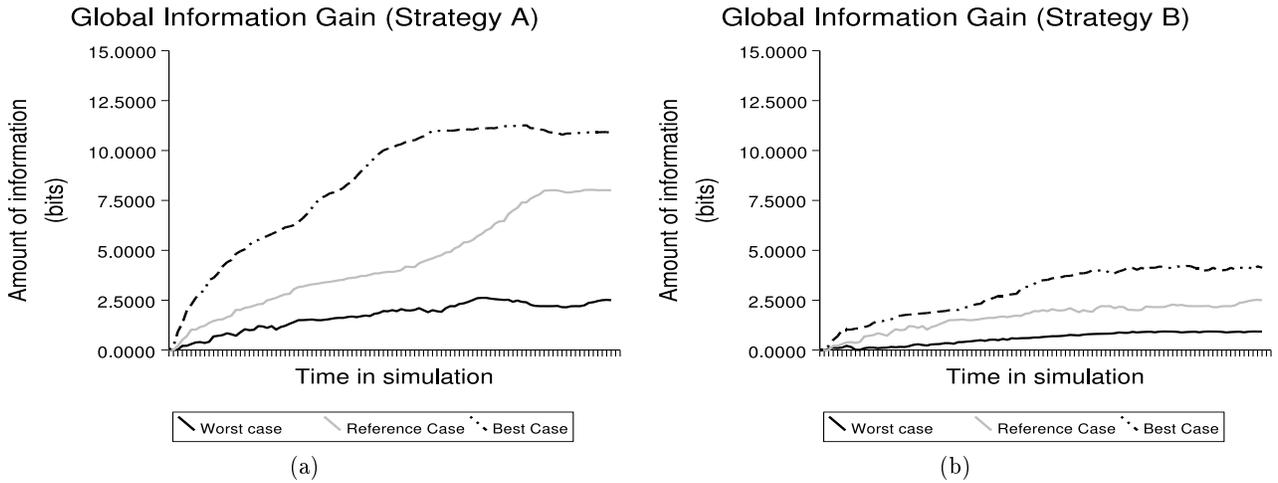
Figure 4: Amount of global information gain

	Worst case					Reference case					Best case			
	D	RA	I_l	I_g		D	RA	I_l	I_g		D	RA	I_l	I_g
SQ	0.0651	**0.4613**	**0.5341**	**0.6295**		0.2470	**0.5908**	**0.6977**	**0.7122**		0.3532	**0.6114**	**0.9196**	**0.9044**
D		0.0293	0.0747	0.1331			0.1965	0.2359	0.2411			**0.2592**	**0.3458**	**0.3442**
RA			0.2154	0.2648				0.4122	0.4371				**0.6639**	**0.7833**
I_l				**0.9308**					**0.9519**					**0.9662**

Table 3: Correlation matrices : Strategy A

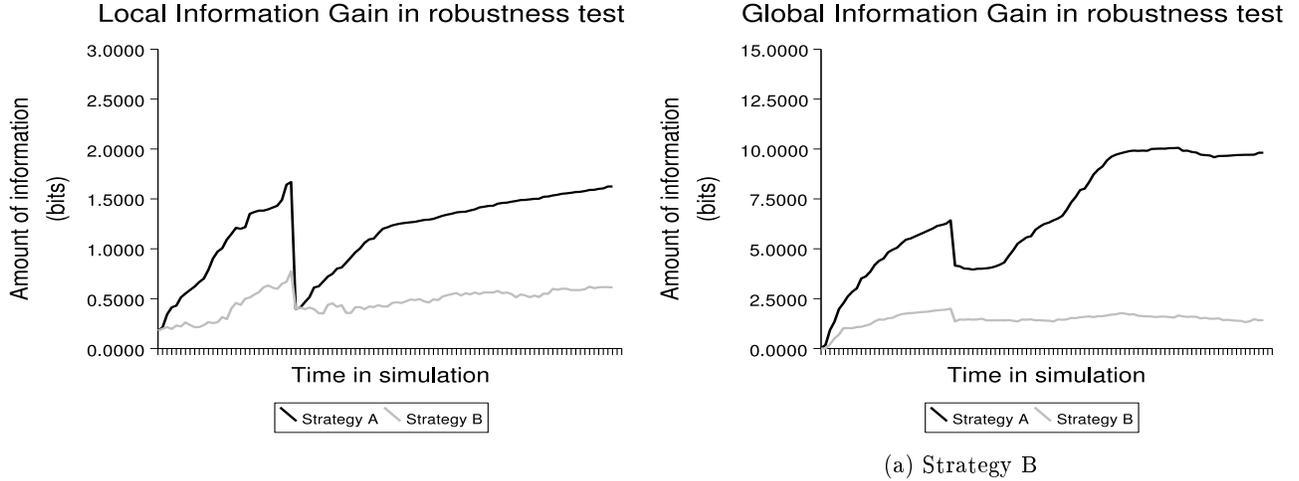|	Worst case					Reference case					Best case			
	D	RA	I_l	I_g		D	RA	I_l	I_g		D	RA	I_l	I_g
SQ	-0.0090	-0.6072	**0.4145**	0.1429		0.0676	-0.5599	**0.3919**	**0.2894**		0.2752	-0.4808	**0.6586**	**0.3982**
D		-0.0012	0.0243	0.0168			0.0016	0.0839	0.0936			0.0001	0.1597	0.2252
RA			0.1156	0.1312				0.2614	0.2351				0.3067	0.1345
I_l				**0.6017**					**0.6599**					**0.6462**

Table 4: Correlation matrices : Strategy B

Figure 5: Amount of information gain in robustness and adaptability test

increases.

Figure 3 and Figure 4 present the amount of local and global information achieved when we use each strategy. Strategy A provides more information than strategy B. Although strategy A attains a maximum of 2.79 bits of local information out of a possible 3 bits, strategy B only accomplished a maximum of 1.19 bits of local information, corresponding to a 40% success. Visually these results show similar trends to the measurements of path formation speed. The highest global information that is achieved when we use strategy A is 11.26 bits of information which indicates an approximately 85% level of success. Notwithstanding the performance of strategy A, only 4.21 bits of global information (approximately 32% successful) is achieved as best when we use strategy B. We note that agents that use strategy B perceive the existence of global information later than agents that use strategy A.

When we add obstacles in the terrain, both local and global information drop. Figure 5 illustrates this scenario. We observe that strategy A self repairs and re-gains the lost information. Strategy B never achieves the pre-obstacle level of information during the remainder of the simulation. Again, these observations are consistent with the levels of emergence observed for these experiments. This apparent relationship is quantified in the next section.

4.3 Correlation

Tables 3 and 4 report correlation statistics achieved between path formation and information measures.

4.3.1 Local information

Path formation speed shows a significant positive correlation to local information in all cases. The correlation is particularly strong in the best case of strategy A. Although correlation is not causative, we conclude in this case at a 1% level of significance that the amount of local information that we achieve can be used to predict path formation speed and vice verse. The scatterplot in figure 6 allows visual assessment of the relationship between path formation speed and local information using the best case. In this case the plot is narrow, indicating strong correlation. Equation

(5) represents the prediction trend line plotted in the scatterplot for these two data sets.

$$y = 0.0646x + 0.0131 \qquad (5)$$

In this equation, y represents the amount of local information and x indicates path formation speed.

Directionality, on the other hand, shows no correlation with the amount of local information, which we attribute to our current approach of measuring directionality at each time-step as a single Boolean value. A weak correlation between robustness and local information exists for all cases, although this is not significant for scenarios with little emergent behaviour.

4.3.2 Global information

The relationship between global information and path formation speed is significant only for the scenarios corresponding to strategy A. We conclude that predicting path formation speed from global information is only possible when strong emergent behaviour is present. A similar trend occurs with respect to global information and robustness.

5. DISCUSSION AND CONCLUSION

This paper investigates correlations between measures of emergence and of information. We present evidence of strong correlations between local/global information and path formation speed and robustness, particularly when clear emergent behaviour is present. This supports our hypothesis that information measures are candidates for predicting the presence of emergent behaviour.

Weaker positive correlations occur in some cases, when emergent behaviour is less distinct. In these cases agent behaviour has a large random component. Random behaviour detracts from emergence, but remains a significant contributor towards information measures. Future work involves refining information measures to reduce this effect.

The contributions of this work are:

- this is a novel way of evaluating the performance of agent based models using information theory.
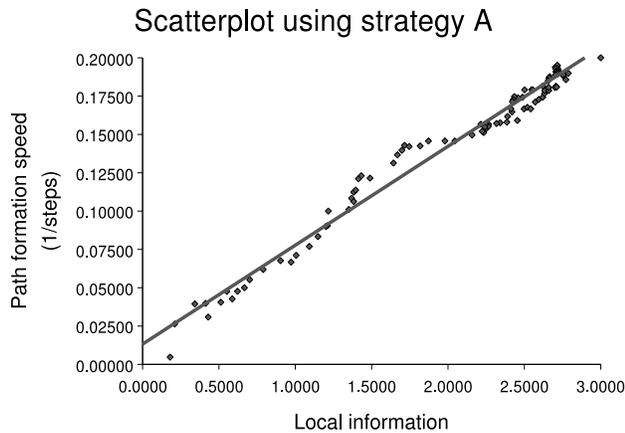
Figure 6: Scatterplot for local information and path formation speed using strategy A

- we provide an evaluation of the relationships between the emergence of the shortest path and the value of stigmergic information in the environment.

- we propose a method for quantifying emergent behaviour.

5.1 Acknowledgement

We would like to thank the entire VRSIG team (2008) at Rhodes University for valuable suggestions and moral support.

6. REFERENCES

[1] Eric Bonabeau, Marco Dorigo, and Guy Theraulaz. *Swarm Intelligence: From Natural to Artificial Systems.* Oxford University Press, 1999.

[2] Stephen F. Bush and Amit B. Kulkarni. Active network management and kolmogorov complexity. In *proceedings of the IEEE OpenArch 2001 Conference,* 2001.

[3] Tom Carter. *An introduction to information theory and entropy.* Complex Systems Summer Schools, June 2006.

[4] Chaomei Chen. An information-theoretic view of visual analytics. *IEEE Computer Graphics and Applications,* 28(1):18–23, 2008.

[5] Chibaya Colin and Bangay Shaun. A probabilistic movement model for shortest path formation in virtual ant-like agents. In *ACM International Conference Proceeding Series; Vol. 226. Proceedings of the 2007 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries,* 2007.

[6] Chibaya Colin and Bangay Shaun. Flock inspired area coverage using wireless boid-like sensor agents. In *proceedings of the 10th International Conference on Computer Modeling and Simulations - EUROSIM/UKSim,* pages 144–149, Los Alamitos, CA, USA, April 2008. IEEE Computer Society.

[7] Garson G. David. *Statnotes: Topics in Multivariate Analysis. Retrieved on 15/08/2008 from*

http://www2.chass.ncsu.edu/garson/pa765/statnote.htm, 2008.

[8] Bavaud F, Chappelier J.C, and Kohlas J. *An introduction to information theory and Applications,* March 2005.

[9] David Feldman. A brief introduction to: Information theory, excess entropy and computational mechanics. College of the Atlantic, ME 04609, October 2002.

[10] Kirsty Kitto. *Modelling and generating complex emergent behaviour.* PhD thesis, The School of Chemistry, Physics and Earth Sciences, The Flinders University of South Australia, June 2006.

[11] David M. Lane. *Computing Pearson's Correlation Coefficient. Retrieved on 15/08/2008 from http://davidmlane.com/hyperstat/A51911.html,* 2007.

[12] Peter Martin. Beginning research on the quantification of spatial order. In *Proceedings of the SIRC 2006- The 18th Annual Colloquium of the Spatial Information Research Center, University of Otago, New Zealand,* November 2006.

[13] Jeremy Miles. *Table of Critical values for Pearson correlation. Retrieved on 15/08/2008 from http://www.jeremymiles.co.uk/misc/tables/pearson.html.* Santa Monica, 2008.

[14] Crutchfield James P. and Feldman David P. Regularities unseen, randomness observed: Levels of entropy convergence. chaos, in press, 2001a. e-print, arxiv.org, cond-mat/0102181. CHAOS, in press, 2001a. E-print, arxiv.org, cond-mat/0102181., February 2001.

[15] Feldman David P. and Crutchfield James P. Discovering noncritical organization: Statistical mechanical, information theoretic, and computational views of patterns in one-dimensional spin systems. santa fe institute working paper., April 1998.

[16] Thomas D Schneider. *Information Theory Primer with an Appendix on Logarithms.* PhD thesis, National Cancer Institute, Laboratory of Experimental and Computational Biology, April 2007.

[17] Richard V. Sole and Octavio Miramontes. Information at the edge of chaos in fluid neural networks. In *Journal Physica D 80 (1995),* volume 80, 1995.

[18] William M.K. Trochim. *Research Methods Knowledge Base: Descriptive statistics: Correlation. Retrieved on 15/08/2008 from http://www.socialresearchmethods.net/kb/statcorr.php,* 2006.

[19] Pere-Pau Vazquez and Mateu Sbert. Perception-based illumination information measurement and light source placement. In *Lecture Notes in Computer Science, Proceedings of ICCS.,* pages 149–157, 2003.