Predicting Cluster Formation in Decentralized Sensor Grids

Astrid Zeman and Mikhail Prokopenko

CSIRO Information and Communication Technology Centre
Locked bag 17, North Ryde, NSW 1670, Australia
mikhail.prokopenko@csiro.au

Abstract. This paper investigates cluster formation in decentralized sensor grids and focusses on predicting when the cluster formation converges to a stable configuration. The traffic volume of inter-agent communications is used, as the underlying time series, to construct a predictor of the convergence time. The predictor is based on the assumption that decentralized cluster formation creates multi-agent chaotic dynamics in the communication space, and estimates irregularity of the communication-volume time series during an initial transient interval. The new predictor, based on the auto-correlation function, is contrasted with the predictor based on the correlation entropy (generalized entropy rate). In terms of predictive power, the auto-correlation function is observed to outperform and be less sensitive to noise in the communication space than the correlation entropy. In addition, the preference of the auto-correlation function over the correlation entropy is found to depend on the synchronous message monitoring method.

1 Introduction

There is a distinction between "Sensor Networks" and "Sensor Grids", as pointed out in recent literature (e.g., [3]): "whereas the design of a sensor network addresses the logical and physical connectivity of the sensors, the focus of constructing a sensor grid is on the issues relating to the data management, computation management, information management and knowledge discovery management associated with the sensors and the data they generate". One significant issue addressed by sensor grids is dynamic sensor-data clustering, aimed at grouping entities with similar characteristics together so that main trends or unusual patterns may be discovered. This is investigated as decentralized clustering in multi-agent Systems [9], dynamic cluster formation in mobile ad hoc networks [7] and decentralized sensor arrays [8, 13, 10]. The latter studies describe dynamic cluster formation as *self-organisation* of dynamic hierarchies, with multiple cluster-heads emerging as a result of inter-agent communications, and indicates that decentralized clustering algorithms deployed in multi-agent systems are "hard to evaluate precisely for the reason of the diminished predictability brought about by self-organisation". The results presented in [13] identified a predictor for the convergence time of dynamic cluster formation, based on the traffic volume of asynchronous inter-agent communications. Following this study, we attempt to adapt a decentralized clustering algorithm to a specific topology (a rectilinear grid) and replace a complicated predictor with a more simple measure, based on synchronized aggregation of multi-agent communications.

Our goal is predicting when the cluster formation will converge to a stable configuration. In achieving this goal, we consider an underlying time series, the traffic volume of inter-agent communications, and relate its irregularity during an initial interval to the eventual convergence time. Clearly, the shorter the initial interval is, the more efficient is the prediction: e.g., when a predicted value exceeds a threshold, agents may adjust parameters and heuristics used in the clustering process.

A simplified version of a decentralized adaptive clustering algorithm developed for evaluation purposes is described in the next section. Section 3 presents the proposed predictor for the convergence time of cluster formation, followed by a discussion of the obtained results.

2 Dynamic Cluster Formation Algorithm

A sensor grid node communicates only with immediate neighbours: all data are processed locally, and only information relevant to other regions of the structure is communicated as a multi-hop message. A cluster-head may be dynamically selected among the set of nodes and become a local coordinator of transmissions within the cluster. Clusters may re-form when new data is obtained on the basis of local sensor signals. Importantly, a cluster formation algorithm should be robust to such changes, failures of individual nodes, communication losses, etc.

As pointed out earlier, our main goal is an analysis of a representative clustering technique in a dynamic and decentralized multi-agent setting, exemplified by a rectilinear sensor grid, *in terms of predictability of its convergence time*. We represent a node sensory reading with a single aggregated value, define “differences” between cells in terms of this value, and cluster nodes while minimizing these “differences”.

The algorithm input is a series of events detected at different times and locations, while the output is a set of non-overlapping clusters, each with a dedicated cluster-head (a network node) and a cluster map of its followers in terms of their sensor-data and relative grid coordinates. The algorithm is described elsewhere [8] and involves a number of inter-agent messages notifying agents about their sensory data, and changes in their relationships and actions. For example, an agent may send a recruit message to another agent, delegate the role of cluster-head to another agent, or declare “independence” by initiating a new cluster. Most of these and similar decisions are based on the clustering heuristic described by Ogston et al. [9], and a dynamic offset range [8]. This heuristic determines if a cluster should be split in two, and the location of this split. Each cluster-head (initially, each agent) broadcasts its *recruit* message periodically, with a broadcasting-period, affecting all agents with values within a particular dynamic offset of the sensor reading detected by this agent. Every *recruit* message contains the sensor-data of all current followers of the cluster-head with their relative coordinates (a cluster map). Under certain conditions, an agent, which is not a follower in any cluster, receiving a *recruit* message becomes a follower, stops broadcasting its own *recruit* messages and sends its information to its new cluster-head indicating its relative coordinates and the sensor reading. However, there are situations when the receiving agent is already a follower in some cluster and cannot accept a recruit message by itself — a recruit disagreement. In this case, this agent *forwards* the received recruiting request to its present cluster-head. Every cluster-head waits for a certain period, collecting all such *forward* messages, at the end of which the clustering heuristic is invoked on the union set of present followers and all agents who *forwarded* their new requests [8, 13]. The cluster-head which invoked the heuristic notifies new cluster-heads about their appointment, and sends their cluster maps to them: a *cluster-information* message.

Here we consider an important variant of this algorithm, obtained by modifying both the message passing mechanism and the message monitoring method. First of all, instead of sending a *forward* message by broadcasting or "flooding" which makes the system quite resilient to noise, we use point-to-point messages incorporating relative grid coordinates, routed through the grid using these coordinates. Secondly, given a reduction in the communication traffic resulting from point-to-point messages, we employ a message monitoring method which allows to more precisely count inter-agent messages for each relative unit of system time. This essentially means that, instead of counting messages asynchronously (separately for each node) and aggregating these amounts for an abstract unit of time, we synchronize the system and precisely aggregate all messages for each time point. This is not always feasible and may incur a high cost, but the expected tradeoff is the simplicity and performance of new predictors.

In addition, using point-to-point messages is a less reliable method, and we specifically introduced errors in the message-passing mechanism, simulating noise in the communication space — in order to verify robustness of the predictors. The new point-to-point messages significantly reduce the communication traffic, without affecting quality (measured by the weighted average cluster diameter [18]) and convergence (measured by the number of times the clustering heuristic was invoked before stability is achieved). While the simulation results show that the algorithm robustly converges and scales well in all cases, the convergence time varies significantly (Figure 1 and Figure 2) — highlighting the need for its better prediction.

The cluster formation is driven by three message types: *recruit*, *cluster-information*, and *forward* messages. The first two types are periodic, while the latter type depends only on the degree of disagreements among cluster-heads. The number of *forward* messages traced in time — the traffic volume of inter-agent communications — provides the underlying time series $\{v(t)\}$ for our predictive analysis.

3 Regularity of multi-agent communication-volume

In this section, we focus on our main objective: prediction of the convergence time T, based on regularity of an initial segment $0, ..., \mathcal{D}$ of the "communication-volume" series $\{v(t)\}$, where $\mathcal{D} < T$ and $v(t)$ is the number of *forward* messages at time t.

It is known that in many experiments, time series often exhibit irregular behavior during an initial interval before finally settling into an asymptotic state which is non-chaotic [1] — in our case, eventually converging to a fixed-point ($v(T) = 0$). The irregular initial part of the series may, nevertheless, contain valuable information: this is particularly true when the underlying dynamics is deterministic and exhibits *transient chaos* [1, 5]. It was conjectured and empirically verified [13] that the described algorithm for dynamic cluster formation creates *multi-agent transient chaotic dynamics*.

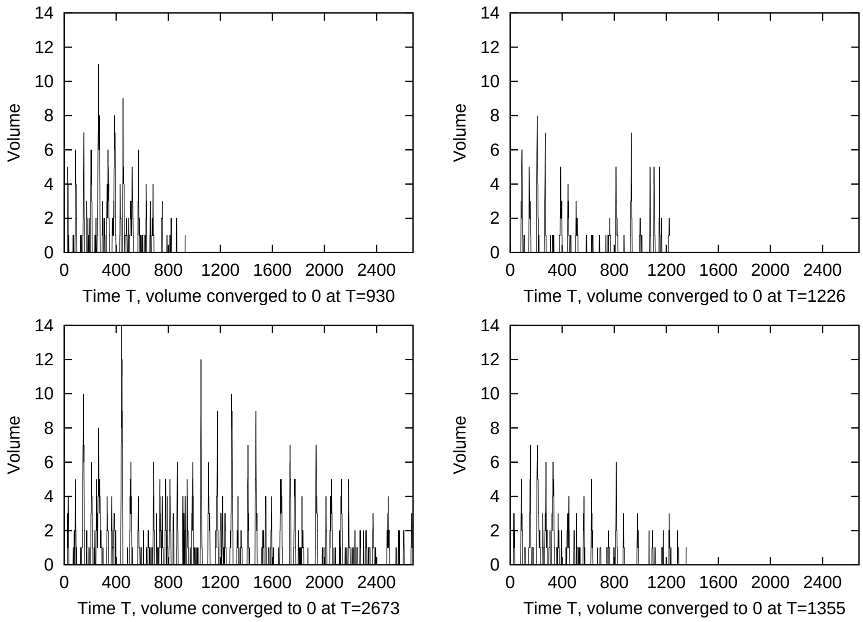

Fig. 1. Varying convergence times T_s for 4 different experiments, $1 \leq s \leq 4$, without noise.

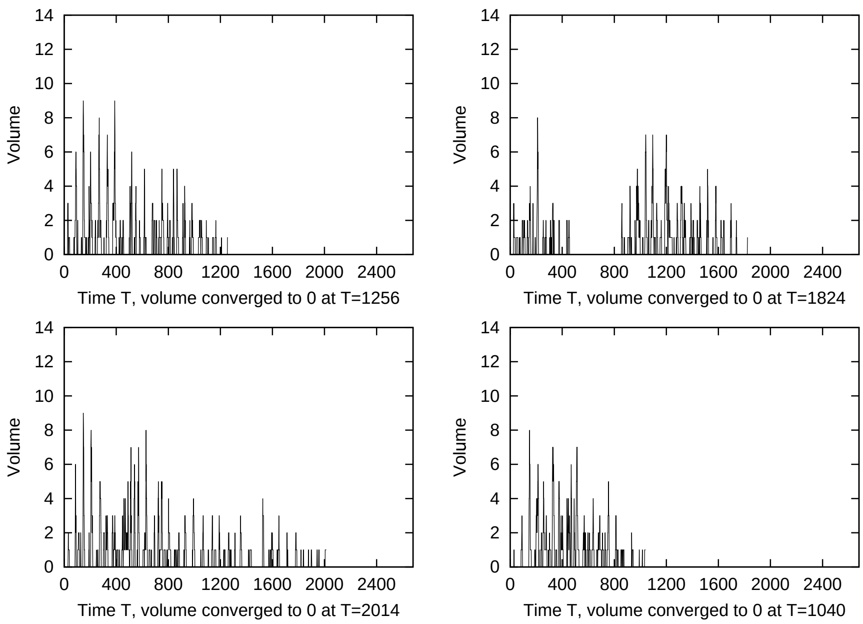

Fig. 2. Varying convergence times T_s for 4 different experiments, with noise.

We intend to follow the same path as the previous study [13], but streamline the predictor estimation by using a simple auto-correlation function as a measure of regularity during the initial interval. For each experiment s, we a) select an initial segment of length $\mathcal{D}$ of the time series; and b) compute the regularity predictor: the auto-correlation function $\gamma(\mathcal{D},\tau)_s$ for a range of integer delays τ:

$$\gamma(\mathcal{D},\tau)_s \;=\; \sum_{t=\tau+1}^{\mathcal{D}} [v_s(t-\tau) - \overline{v_s}]\; [v_s(t) - \overline{v_s}] \;/\; \sum_{t=1}^{\mathcal{D}} [v_s(t) - \overline{v_s}]^2 \;, \qquad (1)$$

where $\overline{v_s}$ is the series mean. Then, c) given the estimates $\gamma(\mathcal{D},\tau)_s$ for all the experiments, correlate them with the observed convergence times T_s by using a linear regression $T = a + b\gamma(\mathcal{D},\tau)$ and the correlation coefficient $\rho(\tau)$ between the series $\{T_s\}$ and $\{\gamma(\mathcal{D},\tau)_s\}$. This would allow us to predict the time T_s of convergence to $v_s(T_s) = 0$, as $T_s = a + b\gamma(\mathcal{D},\hat{\tau})_s$, for the delay $\hat{\tau}$ providing the best fit: the maximum of $\rho(\tau)$.

The auto-correlation is obviously limited to measuring only linear dependencies, and the study [13] considered a more general and elaborate approach, based on the Kolmogorov-Sinai entropy K, also known as metric entropy [6, 16], and its generalization to the order-q Rényi entropy K_q [15]. The entropy K or K_q is an entropy per unit time, or an "entropy rate", and is a measure for the rate at which information about the state of the system is lost in the course of time. In particular, the predictor estimated the "correlation entropy" K_2 using Grassberger and Procaccia algorithm [4]. The predictor based on K_2 uses the initial segment of length $\mathcal{D}$ of the observed time series $\{v(t)\}$ in "converting" or "reconstructing" the dynamical information in one-dimensional data to spatial information in the τ-dimensional embedding space [17], and also depends on the length $\mathcal{D}$ and the embedding dimension τ.

The auto-correlation function $\gamma(\mathcal{D},\tau)$, equation (1), was reported to be not sufficient for predictive purposes: the highest correlation coefficient $\rho(\tau)$ between convergence times T_s and auto-correlations $\gamma(\mathcal{D},\tau)_s$, for a range of delays τ, was only 0.52, while the predictor based on the entropy K_2 attained the maximum $\rho = 0.90$. In the following section we shall contrast these two measures, $\gamma(\mathcal{D},\tau)$ and $K_2(\mathcal{D})$, for the new communication and monitoring mechanisms, with and without noise.

4 Experimental Results

The experiments included three scenarios: (i) noiseless communications; (ii) 1% loss of messages; and (iii) 2% loss of messages. Each scenario included 20 runs of the clustering algorithm on an 8×8 grid with 50 events, tracing the communication-volume time series $\{v(t)\}$. We then selected an initial segment $\mathcal{D} = 800$, and carried out the steps b) and c) described in the previous section. Given data of $s = 1, ..., 20$ experiments: the 2-dimensional array $\gamma(\mathcal{D},\tau)_s$ for varying τ and each s, the correlation coefficient $\rho(\{T_s\}, \{\gamma(\mathcal{D},\tau)_s\})$ was determined for the range of τ, based on the auto-correlation predictor $\gamma(\mathcal{D},\tau)$. The data are plotted in Figure 3 for the scenarios (i), (ii) and (iii). The corresponding maximum values of $\rho(\hat{\tau})$ degrade with noise as expected: from (i) $\rho(8) = 0.98$ to (ii) $\rho(23) = 0.82$ to (iii) $\rho(91) = 0.69$. As the level of noise grows, the maximums are attained at increasing delays τ: (i) $\hat{\tau} = 8$; (ii) $\hat{\tau} = 23$; and (iii) $\hat{\tau} = 91$.

At the same time, the predictor based on $K_2(\mathcal{D})$ was sensitive to the higher noise levels: the best obtained correlation values were: (i) $\rho(71) = 0.83$, (ii) $\rho(89) = 0.83$, and (iii) $\rho(72) = 0.37$, as shown in Figure 4. Without noise it performed as expected, maintained the performance under 1% loss of messages, but the increase in the noise by an extra percent resulted in more than 50% loss in predictive power. This can simply be explained by the fact that the extra noise made the underlying dynamics unstructured in the phase space created by the considered embedding dimensions [13], and this can be recovered by increasing their number. Nevertheless, from a practical point of view, it is rarely feasible, and the alternative predictor based on a simple auto-correlation function, is preferable as it is less sensitive to noise in the communication space.

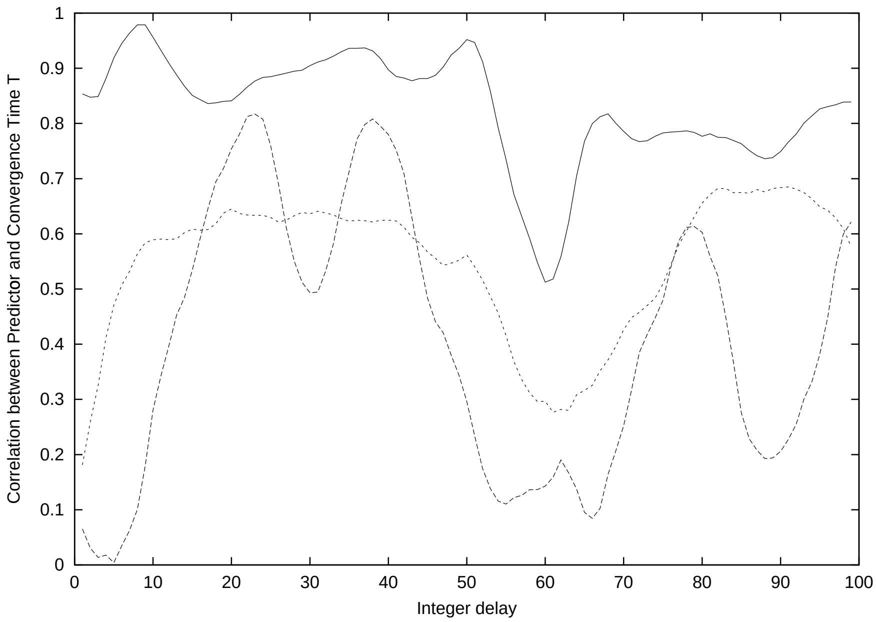

Fig. 3. The correlation coefficient ρ between the series $\{T_s\}$ and predictor $\{\gamma(\mathcal{D}, \tau)_s\}$, for the scenarios (i) solid lines, (ii) dashed lines, and (iii) dotted lines.

We would like to point out that the noise in communication space (missed messages) considered in this paper should be distinguished from the noise in the traffic monitoring method created by its own asynchrony. The preference of the auto-correlation function over the correlation entropy, as the convergence time predictor, is conditional on the synchronous message monitoring method. If the underlying communication traffic is estimated asynchronously, then the observations reported in [13] indicate that the correlation entropy is preferred to the auto-correlation function (even in the presence of noise in the communication space). The reason for this difference is the calculation of correlations: the auto-correlation function simply “matches” separate time points (therefore, it is sensitive to shifts in the time series brought about by asynchronous monitoring), while the correlation entropy “matches” patterns or templates in the time series and is, hence, resilient to possible shifts due to asynchronous monitoring.

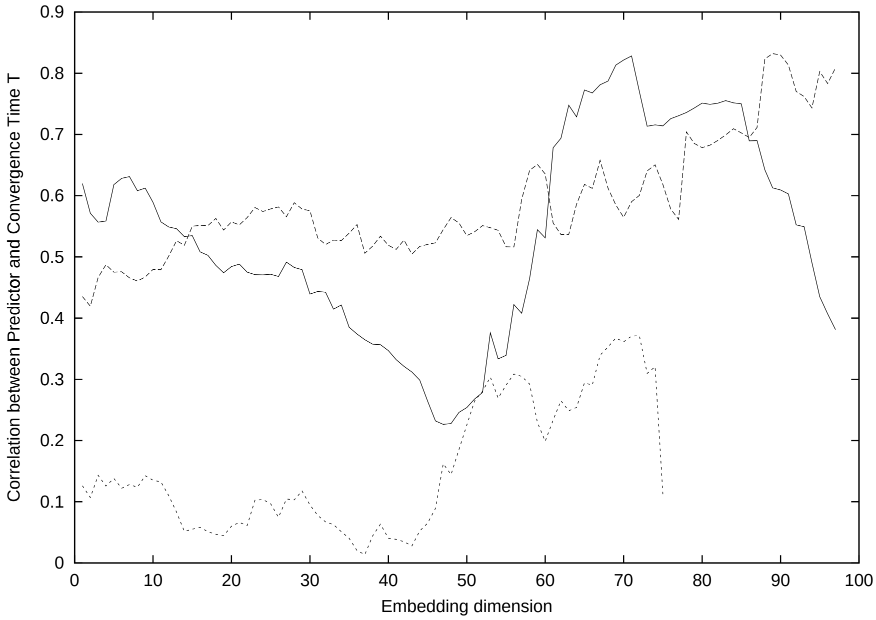

Fig. 4. The correlation coefficient ρ between the series $\{T_s\}$ and predictor $\{K_2(\mathcal{D})_s\}$, for the scenarios (i) solid lines, (ii) dashed lines, and (iii) dotted lines (the last scenario did not have a sufficiently long time series to embed in higher dimensions).

5 Conclusions

We considered decentralized and dynamic cluster formation in multi-agent sensor grids, proposed and experimentally evaluated a new predictor for the convergence time of cluster formation. The new predictor, based on the auto-correlation function $\gamma(\mathcal{D}, \tau)$, was contrasted with the predictor $K_2(\mathcal{D})$ based on the generalized correlation entropy of the volume of the inter-agent communications [13].

The results indicate that either predictor can be well correlated with the time of cluster formation. However, their applicability depends on the type of the communication traffic's monitoring: if the employed measure is asynchronous then $K_2(\mathcal{D})$ is preferred, otherwise, if messages can be aggregated synchronously, the auto-correlation function $\gamma(\mathcal{D}, \tau)$ should be preferred. In addition, the correlation entropy $K_2(\mathcal{D})$ was shown to be adversely affected, as a predictor, by noise in the communication space.

Efficient and reliable algorithms for cluster formation in sensor grids may include a convergence predictor as a feedback to the algorithms. Such predictors are unlikely to implement measures with a global view, when full information on nodes' states and their inter-connections is available. Instead, a more promising approach is to develop measures that can work with partial information, obtained locally: *localizable* measures [14, 11, 12, 2]. The analysis and results presented here and in [13] make a step towards localizable measures defined on the inter-agent communication space, and highlights their applicability in decentralized, dynamic and asynchronous sensor grids. Another direction of future research is scale-free sensor grids.

Acknowledgements. The authors are grateful to Piraveenan Mahendra and Peter Wang for helpful discussions and valuable contributions to earlier versions of the algorithm.

References

1. Dhamala, M., Lai, Y.C., Kostelich, E.J. Analyses of transient chaotic time series. *Physical Review E*, 64, 056207, 1–9, 2001.
2. Foreman, M., Prokopenko, M., Wang, P. Phase Transitions in Self-organising Sensor Networks. Banzhaf, W., Christaller, T., Dittrich, P., Kim, J.T. Ziegler, J. (Eds.) *Advances in Artificial Life - Proceedings of the 7th European Conference on Artificial Life*, 781–791, LNAI 2801, Springer, 2003.
3. Sensor Grid for Air Pollution Monitoring, 2004, Ghanem, M., Guo, Y., Hassard, J., Osmond, M., and Richards, M. In Proceedings of the 3rd UK e-Science All-hands Conference AHM 2004, 106–113, Nottingham UK, 2004.
4. Grassberger, P. and Procaccia, I. Estimation of the Kolmogorov entropy from a chaotic signal. *Physical Review A*, 28(4):2591, 1983.
5. Jánosi, I.M., and Tél, T. Time series analysis of transient chaos. *Physical Review E*, 49(4):2756–2763, 1994.
6. Kolmogorov, A.N. Entropy per unit time as a metric invariant of automorphisms. *Doklady Akademii Nauk SSSR*, 124:754–755 (Russian), 1959.
7. Lin R., and Gerla, M. Adaptive Clustering for Mobile Wireless Networks. *IEEE Journal on Selected Areas in Communications*, 1265–1275, September 1997.
8. Mahendra, P., Prokopenko, M., Wang, P., Price, D.C. Towards Adaptive Clustering in Self-monitoring Multi-Agent Networks. In R. Khosla, R. J. Howlett, L. C. Jain (eds.) *Knowledge-Based Intelligent Information and Engineering Systems, 9th International Conference, KES 2005, Melbourne, Australia, September 14-16, 2005, Proceedings, Part II*, Lecture Notes in Computer Science, Vol. 3682, 796–805, 2005.
9. Ogston E., Overeinder, B., Van Steen, M., and Brazier, F. A Method for Decentralized Clustering in Large Multi-Agent Systems. The 2nd International Joint Conference on Autonomous Agent and Multi Agent Systems, 798–796, 2003.
10. Olsson, L., Nehaniv, C. L., Polani, D., Sensory Channel Grouping and Structure from Uninterpreted Sensor Data, 2004 NASA/DoD Conference on Evolvable Hardware (EH'04), 153-160, 2004.
11. Prokopenko, M., Wang, P., Price, D.C., Valencia, P., Foreman, M., Farmer, A.J. Self-organising Hierarchies in Sensor and Communication Networks. *Artificial Life*, Special issue on Dynamic Hierarchies, Vol. 11(4), 407–426, 2005.
12. Prokopenko, M., Wang, P., Foreman, M., Valencia, P., Price, D., Poulton, G. On connectivity of reconfigurable impact networks in ageless aerospace vehicles. *Journal of Robotics and Autonomous Systems*, Vol. 53, 36–58, 2005.
13. Prokopenko, M., Mahendra, P., Wang, P. On Convergence of Dynamic Cluster Formation in Multi-Agent Networks. In M. S. Capcarrère, A. A. Freitas, P. J. Bentley, C. G. Johnson, J. Timmis (eds.) *Advances in Artificial Life, 8th European Conference, ECAL 2005, Canterbury, UK, September 5-9, 2005, Proceedings*, Lecture Notes in Computer Science, Vol. 3630, 884–894, 2005.
14. Prokopenko, M., Wang, P., Price, D. Complexity Metrics for Self-monitoring Impact Sensing Networks, Proceedings of 2005 NASA/DoD Conference on Evolvable Hardware (EH-05), Washington D.C., USA, 2005.
15. Rényi, A. *Probability theory*. North-Holland, Amsterdam, 1970.
16. Sinai, Ya.G. On the concept of entropy of a dynamical system. *Doklady Akademii Nauk SSSR*, 124:768-771 (Russian), 1959.
17. Takens, F. Detecting strange attractors in turbulence. *Dynamical systems and turbulence*, LNM Vol.898, Springer, Berlin, 1981.
18. Zhang, T., Ramakrishnan, R., Livny, M. BIRCH:A New Data Clustering Algorithm and Its Applications. *Data Mining and Knowledge Discovery*, 1(2), 141-182, 1997.