# Automatic Shape Independent Clustering Inspired by Ant Dynamics

**Prithwish Chakraborty ,Gourab Ghosh Roy,**

**Spandan Sinha, Soumya Bose, Ankur Mondal, and Swagatam Das**

Dept. of Electronics and Telecommunication Engineering
Jadavpur University, Kolkata, India

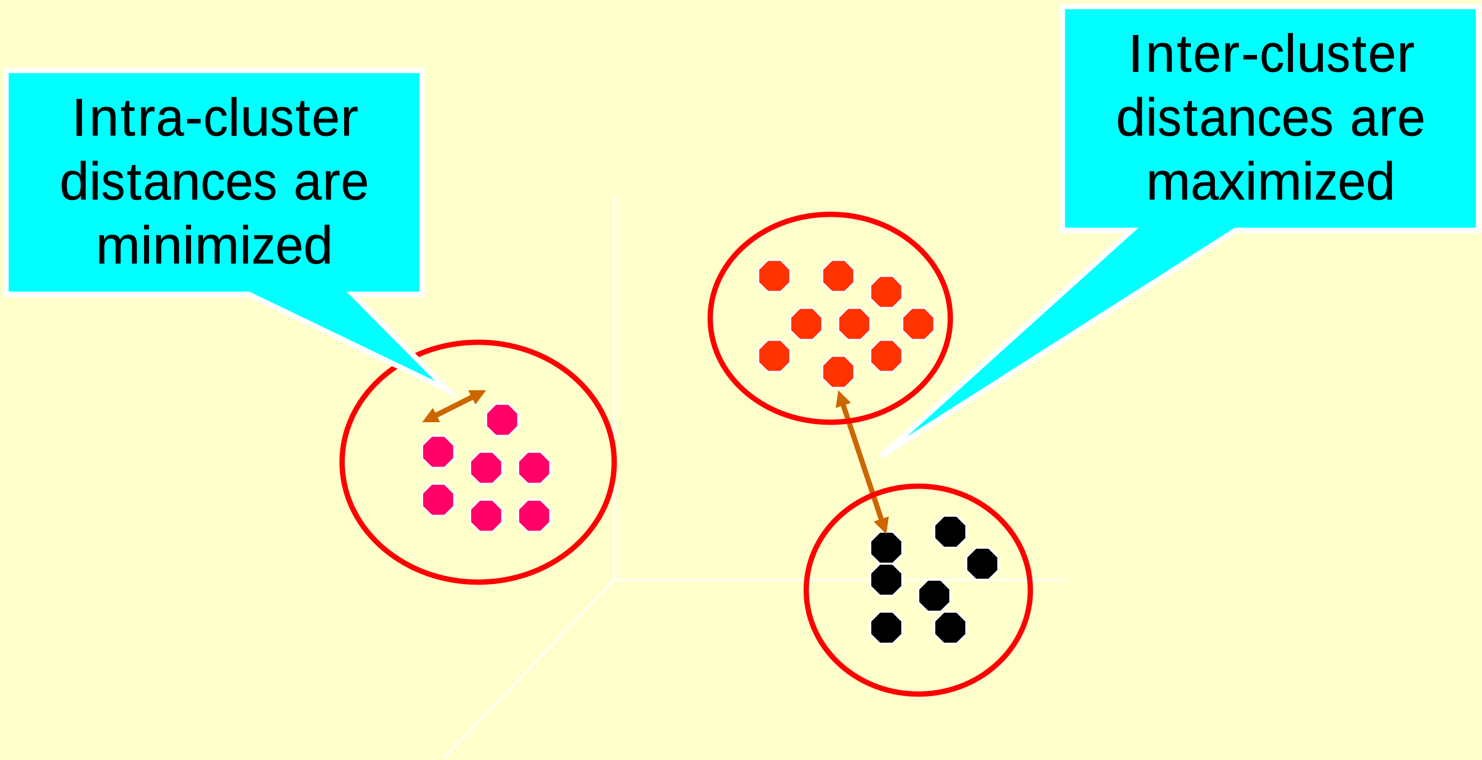
prithwish1611@gmail.com , myself_gourab@yahoo.co.in,
spandan0@gmail.com,1988.soumya@gmail.com, mondal.ankur@gmail.com,
swagatamdas19@yahoo.co.in

# Things to be Addressed

- **Clustering**
- **Traditional clustering algorithms and their drawbacks.**
- **Inspiration drawn from Ant dynamics**
- **Illustration of the algorithm**
- **Experimental results**

# What is Cluster Analysis?

- Finding groups of objects such that the objects in a group will be similar (or related) to one another and different from (or unrelated to) the objects in other groups



# Classification of clustering algorithms

- Sequential
- Hierarchical
- Partitive and Cost function optimization
- Graph-Theoretic
- Information Theory-based

# Some Most Popular Algorithms

- K-means algorithm
- Fuzzy C-means algorithm
- Hierarchical Single-link Agglomerative
- Hierarchical Average-link Agglomerative
- Hierarchical Complete-link Agglomerative
- Genetic Clustering Algorithms (Based on Evolutionary Computing Techniques)

# Disadvantages of available methods

- Some algorithms needs the number of clusters to be pre-specified.
- Different prototypes required for  differently shaped and especially non-globular clusters by some algorithms.
- Increased time-complexity to design efficient and robust algorithms e.g. GA – not suitable for real time applications.
- Some algorithm face difficulty when data contains outliers and when the clusters are of different sizes, densities and non-globular shapes.

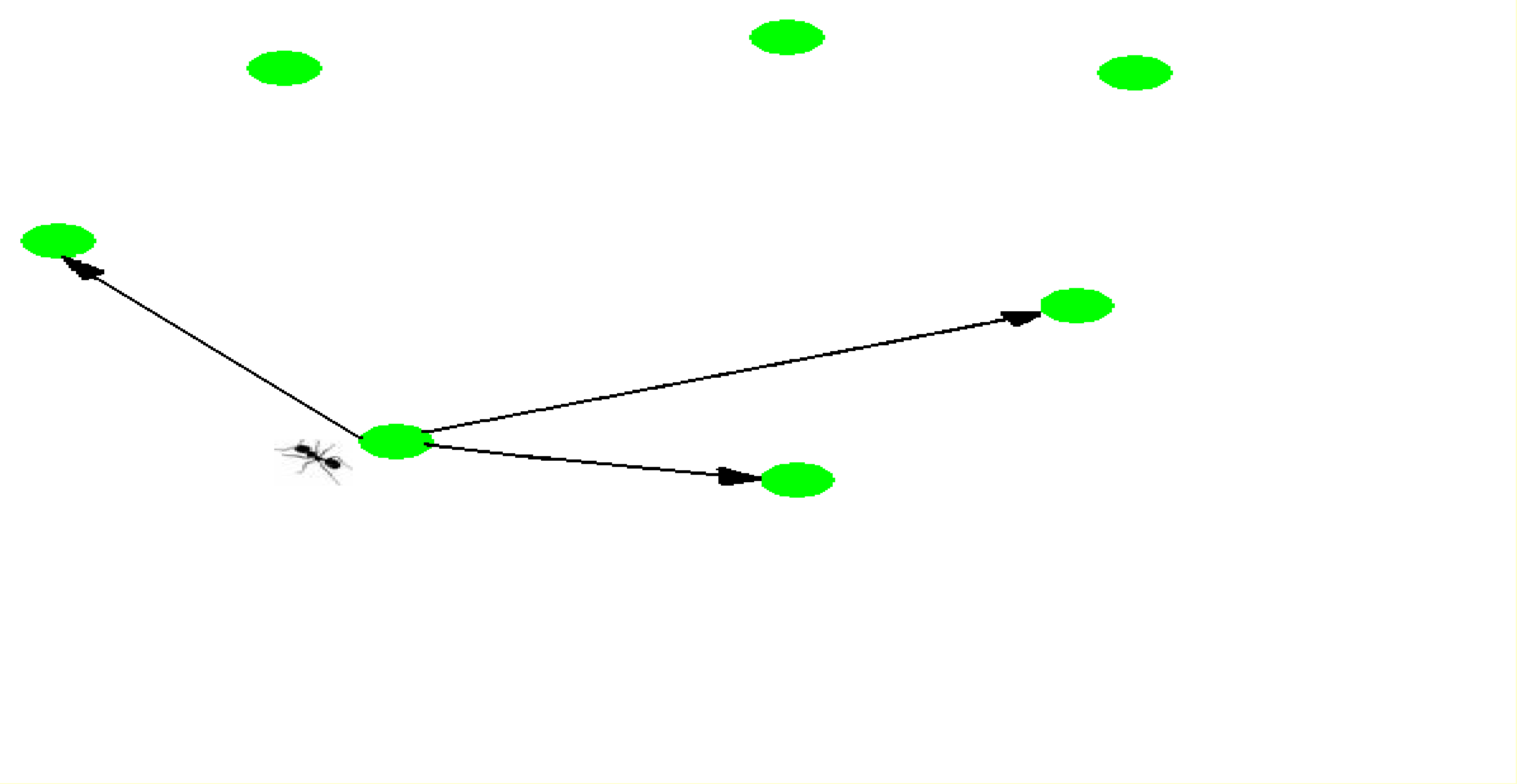
# Some aspects of ant dynamics

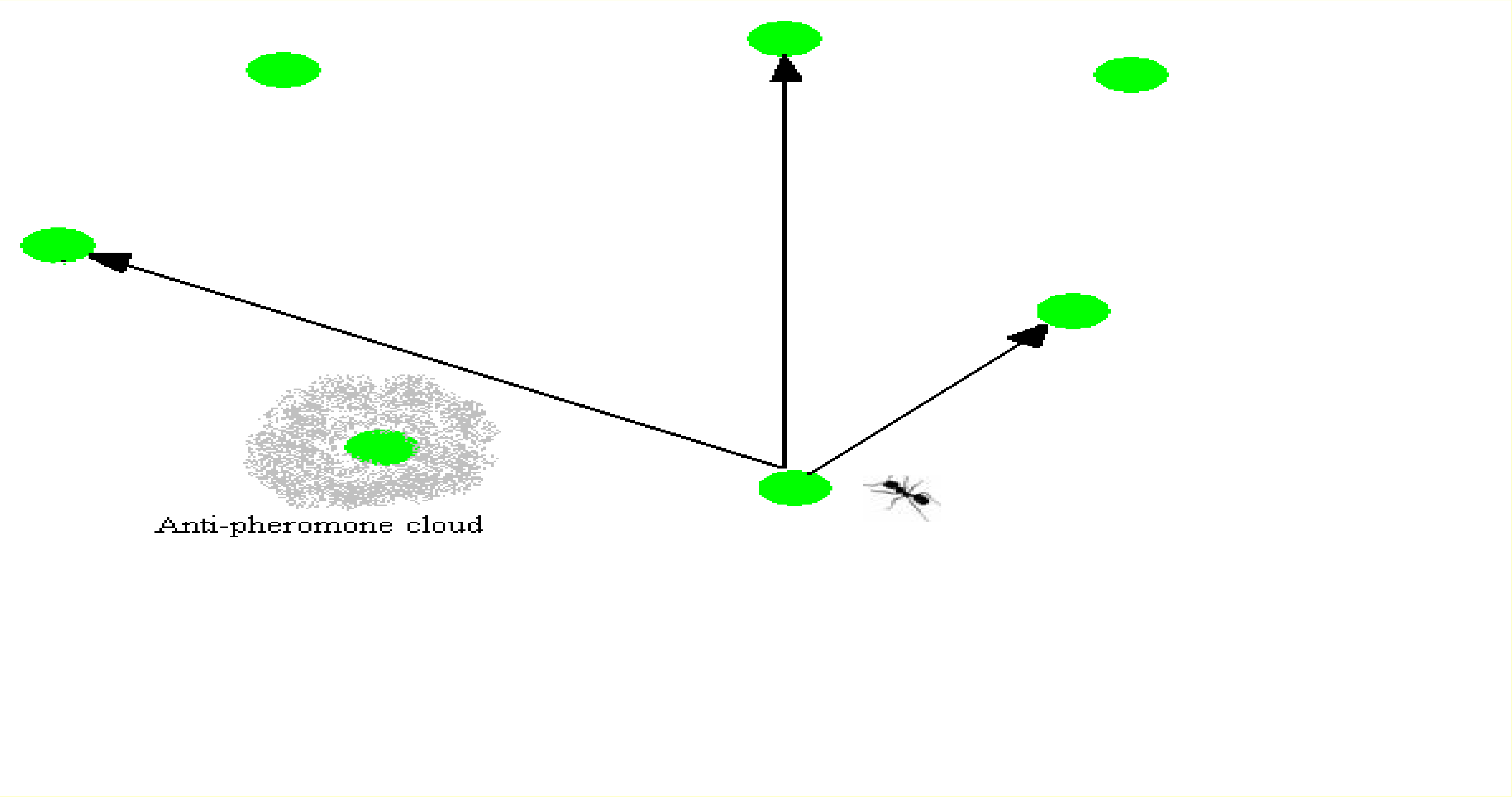
- Ants deposit a chemical substance called pheromone over the path it moves.
- Ants are attracted to a location depending on the amount of pheromone present there.
- Pheromone deposited at a point decays with time.

# Inspiration drawn from ant dynamics – concept of anti-pheromone

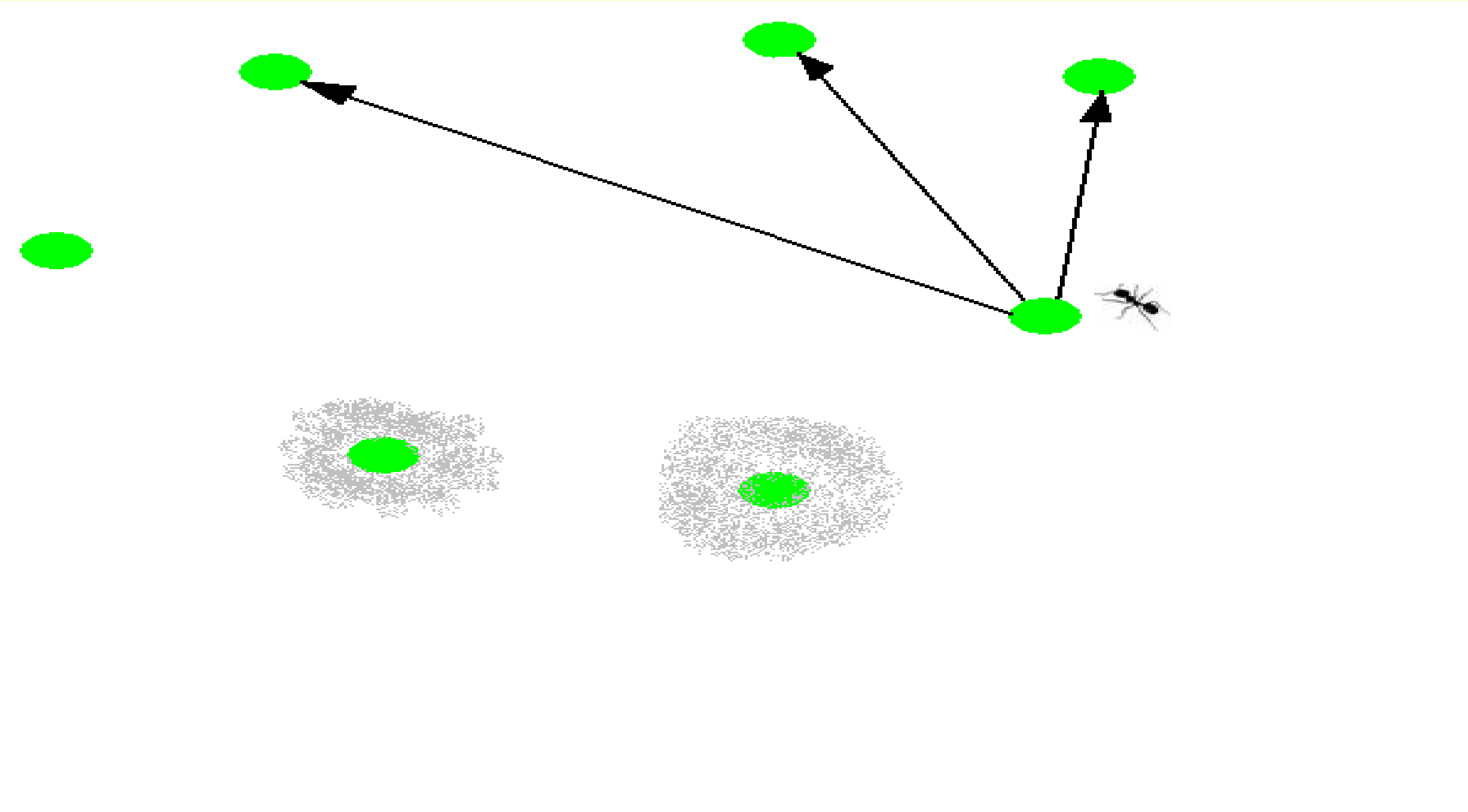
- Artificial-ants are designed to scan the data space.
- They deposit anti-pheromone , a substance which repels the pseudo ant; over its path.
- Anti-pheromone decays with time.

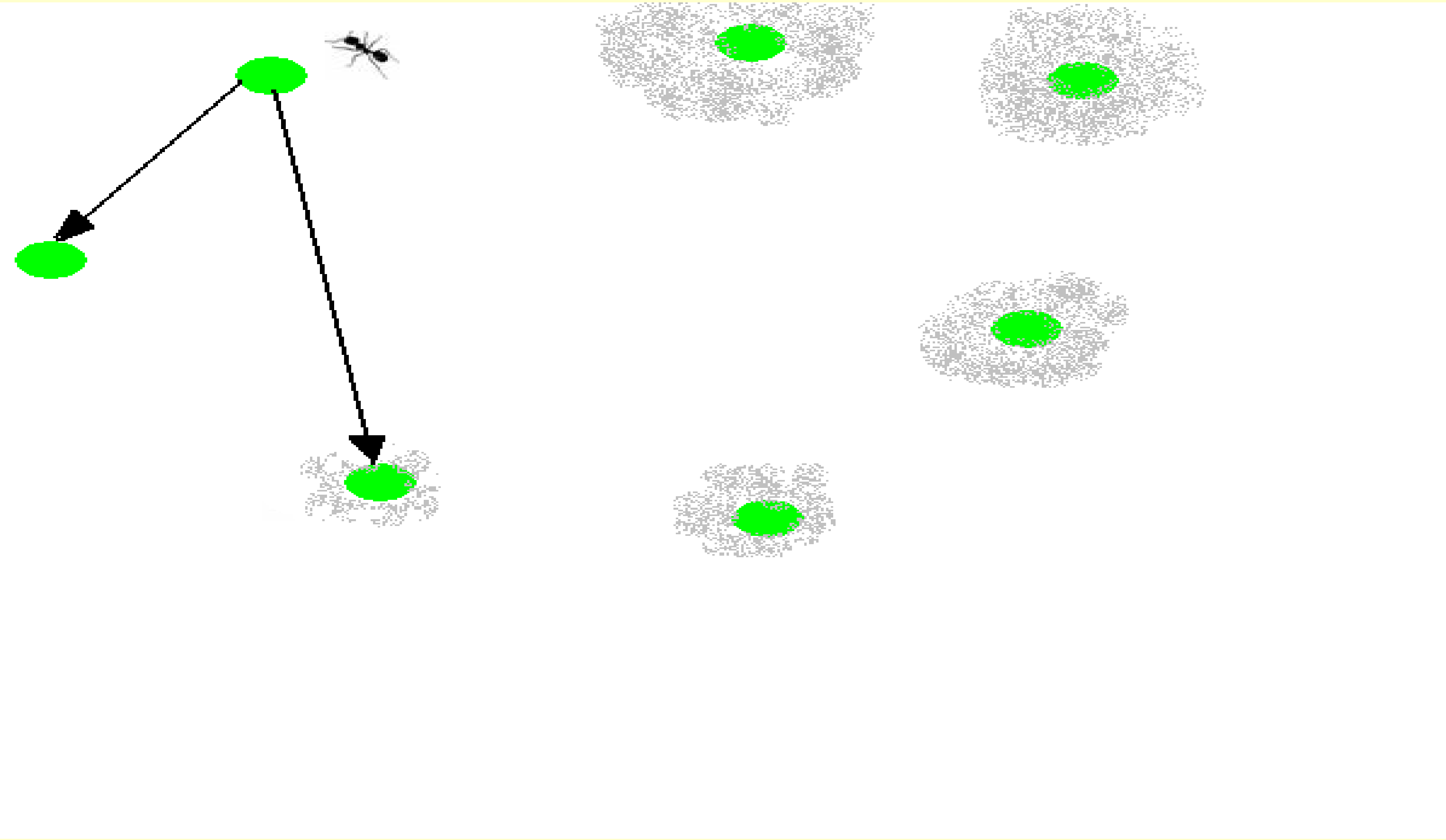
# Algorithm Illustration






Anti-pheromone deposit





# Experimental results

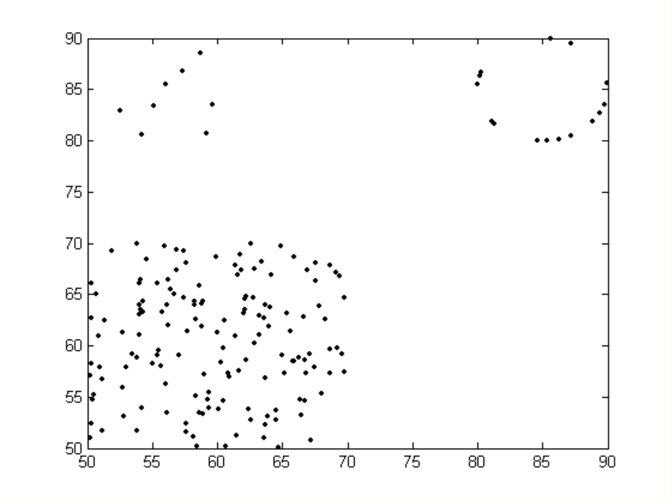
- Parameters for ant inspired clustering are estimated roughly as follows:-

$$penalty = 100 * \max(dist);$$

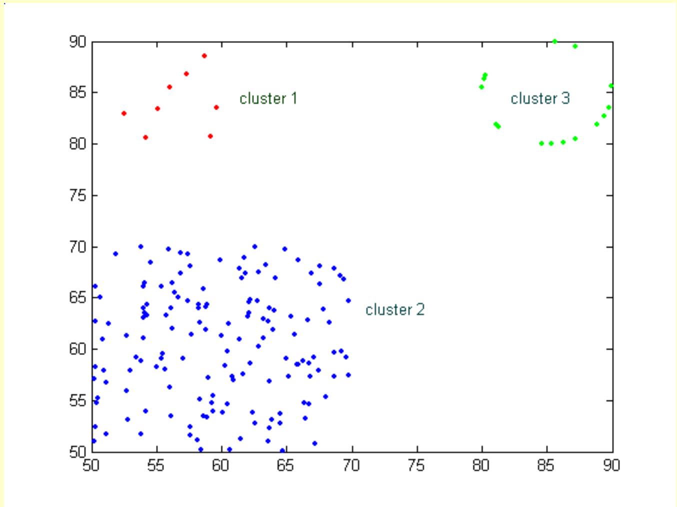
$$factor = 0.15 * \mathrm{Std.}(dist);$$

$$jump = 10 * \{avg(dist) + 3 * s.d(dist)\}$$

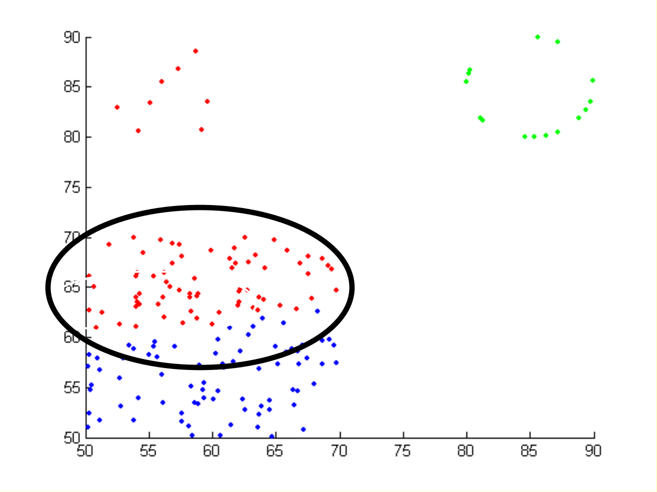
# Clustering results for synthetic data set 1



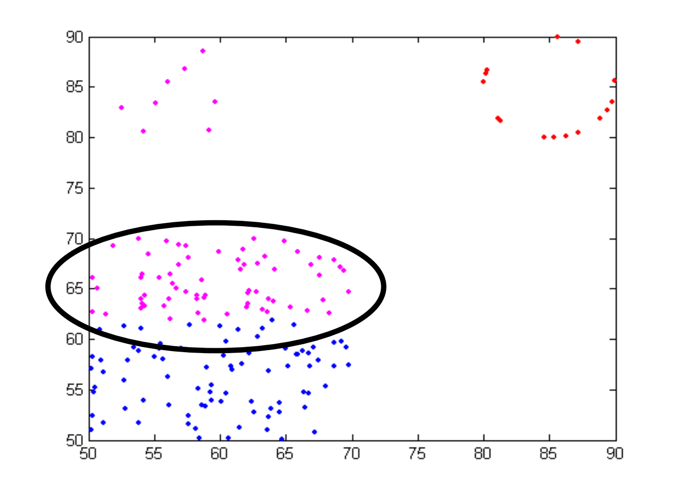

**Unlabelled Synthetic Data_1**




**Ant inspired clustering**

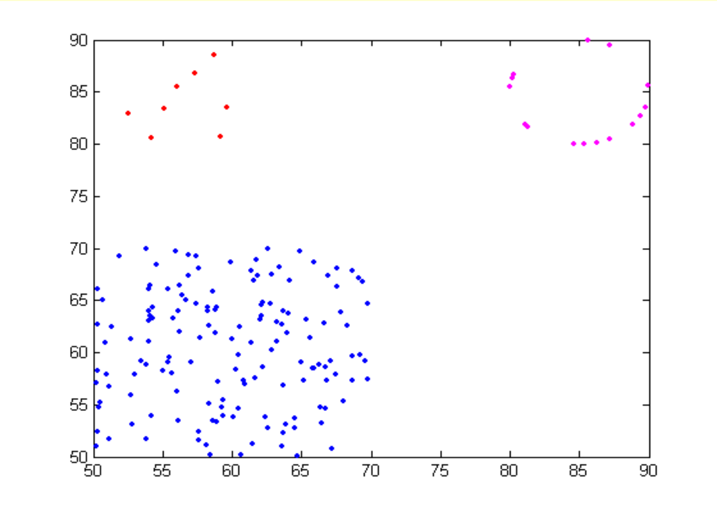


**K means**

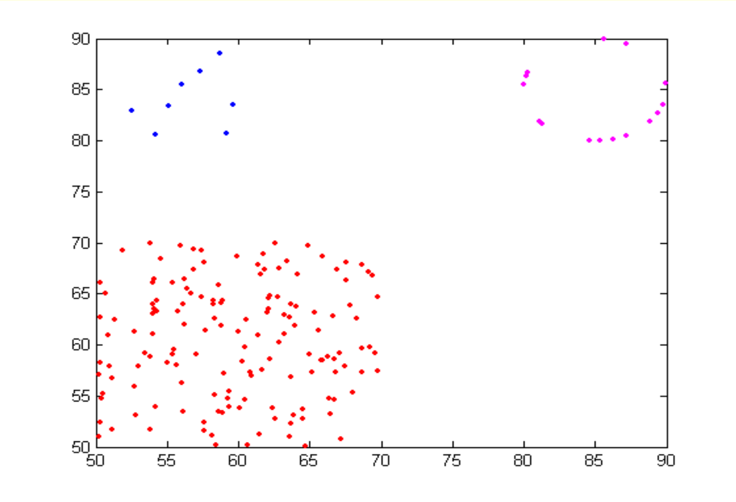


**FCM**

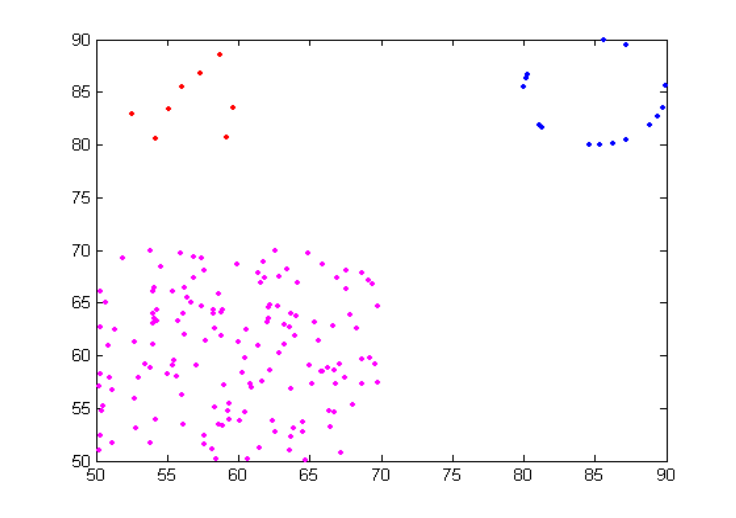
# Clustering results for synthetic data set 1(contd …)




**Hierarchical- average link**

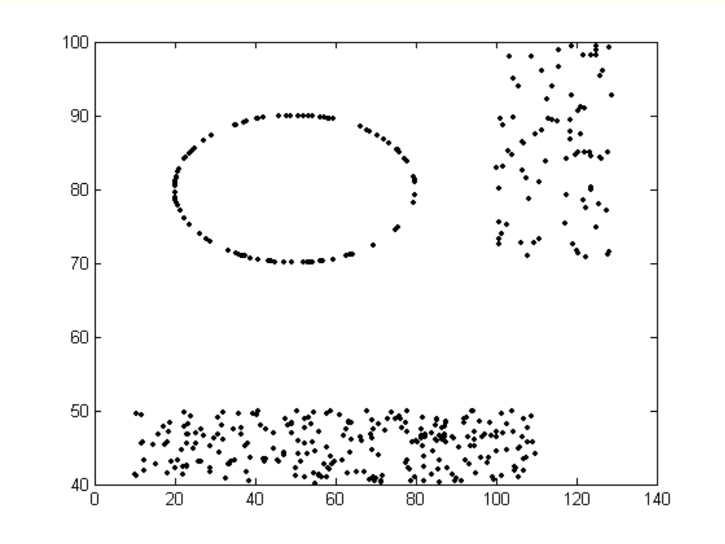



**Hierarchical- single link**

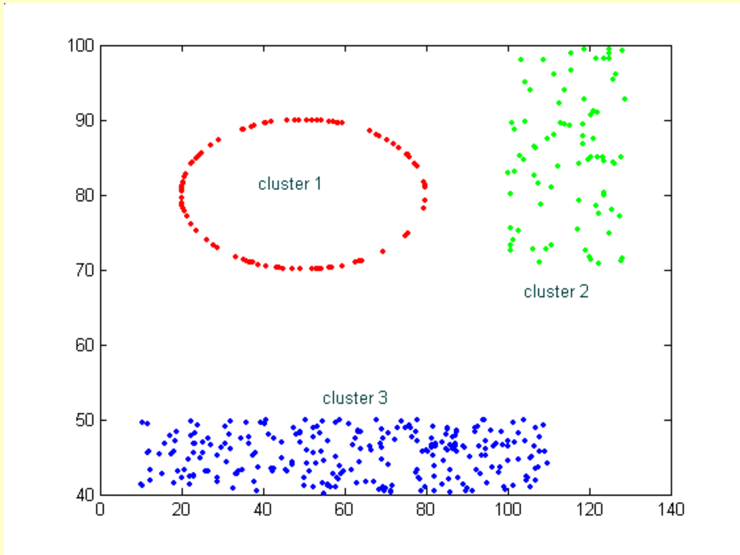



**Hierarchical- complete link**

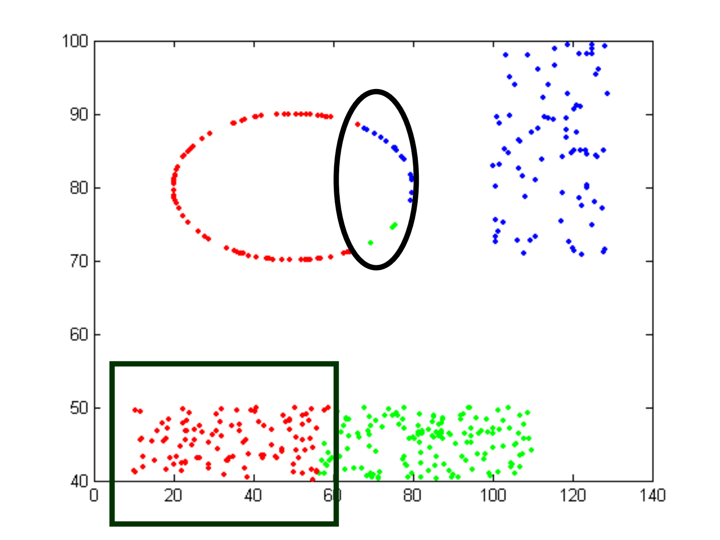
# Clustering results for synthetic data set 2



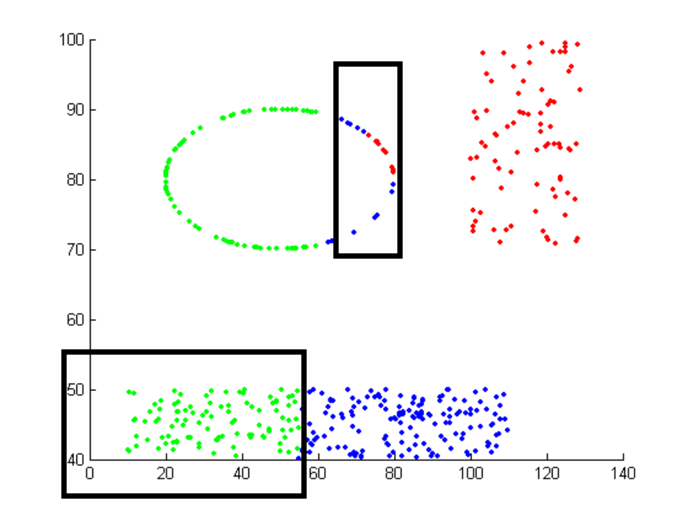

Un-clustered




Ant inspired clustering

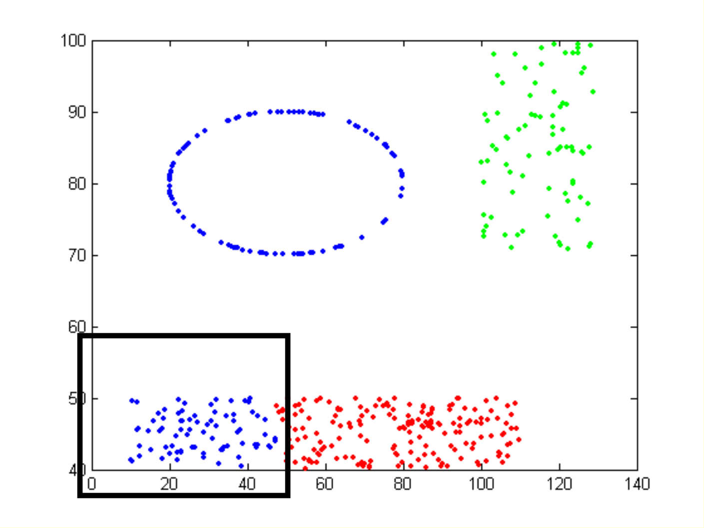



K means

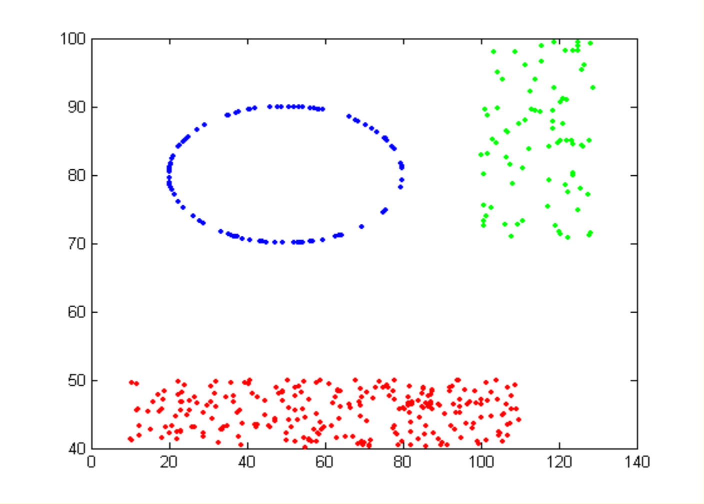



FCM

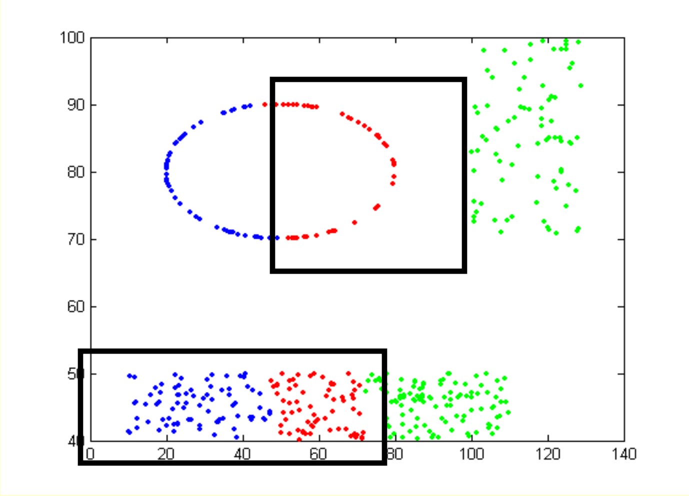
# Clustering results for synthetic data set 2(contd…)




**Hierarchical- average link**

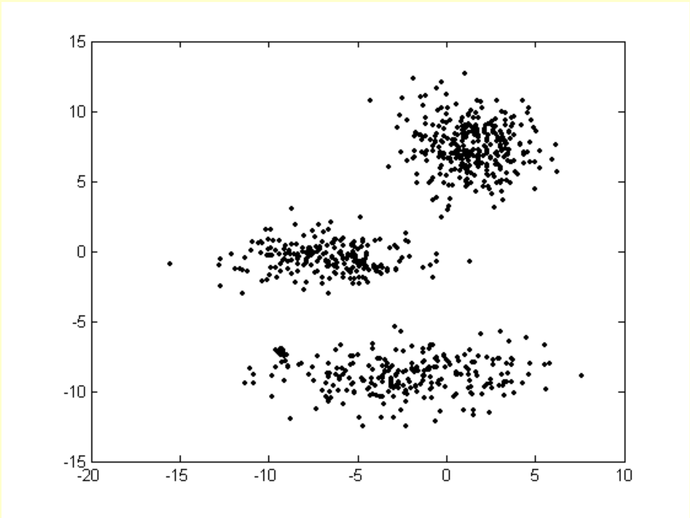



**Hierarchical- single link**

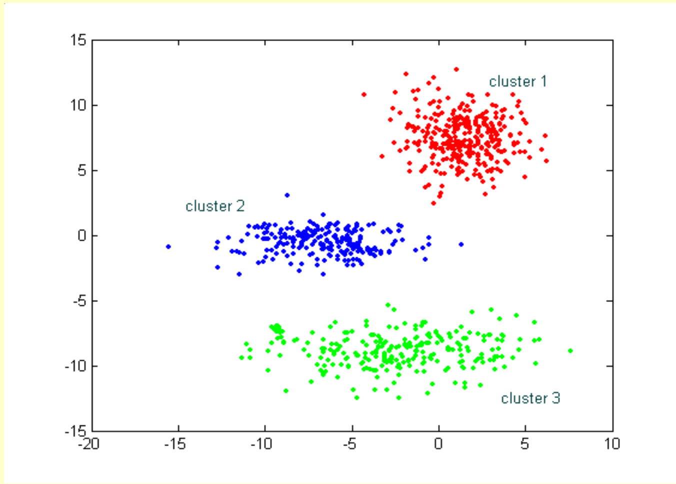



**Hierarchical- complete link**

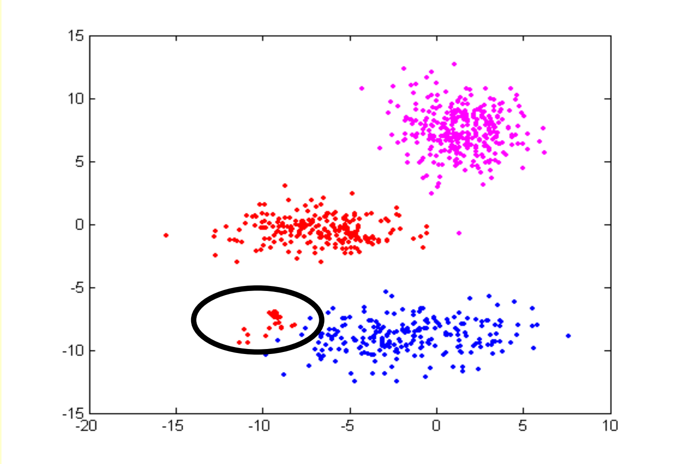
# Clustering results for synthetic data set 3



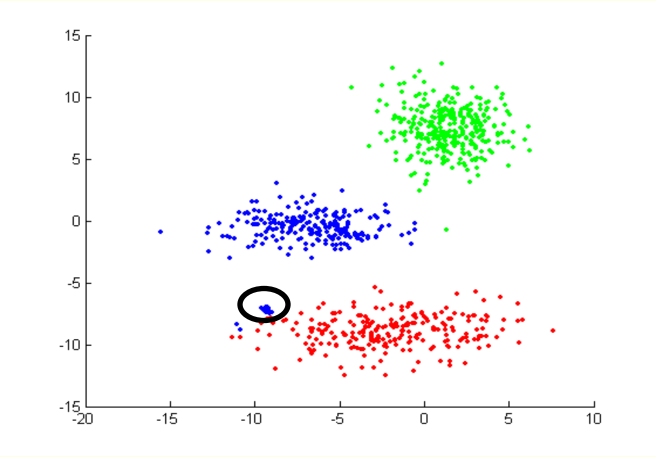

**Un-clustered**




**Ant inspired clustering**

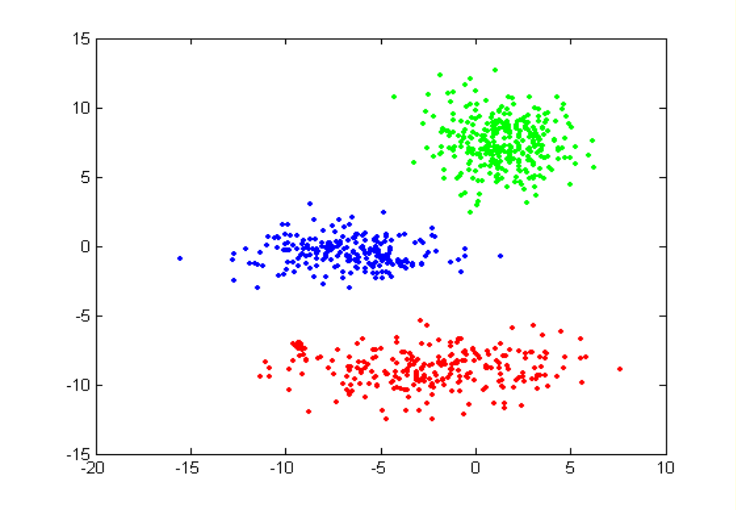



**K means**

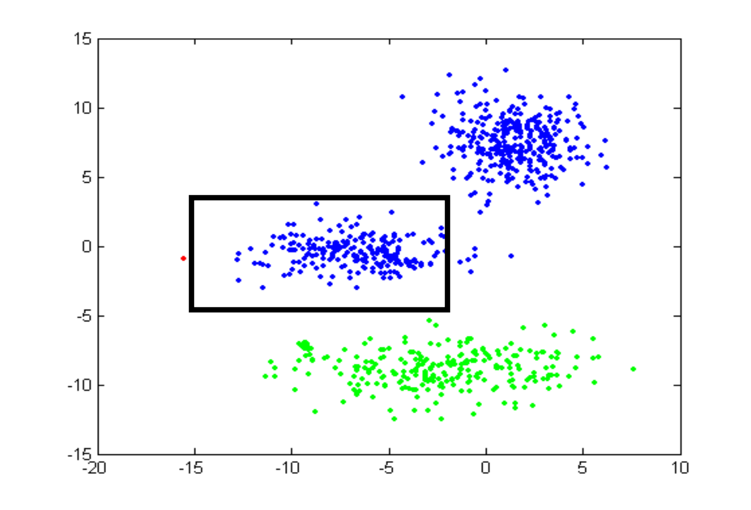



**FCM**

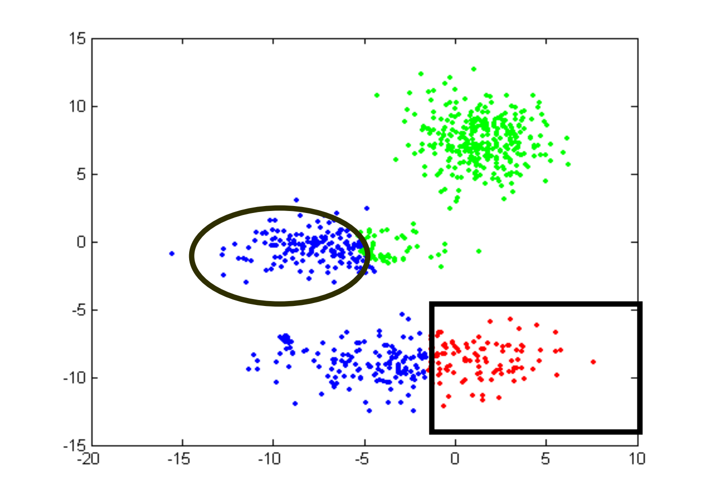
# Clustering results for synthetic data set 3 (contd….)




**Hierarchical- average link**

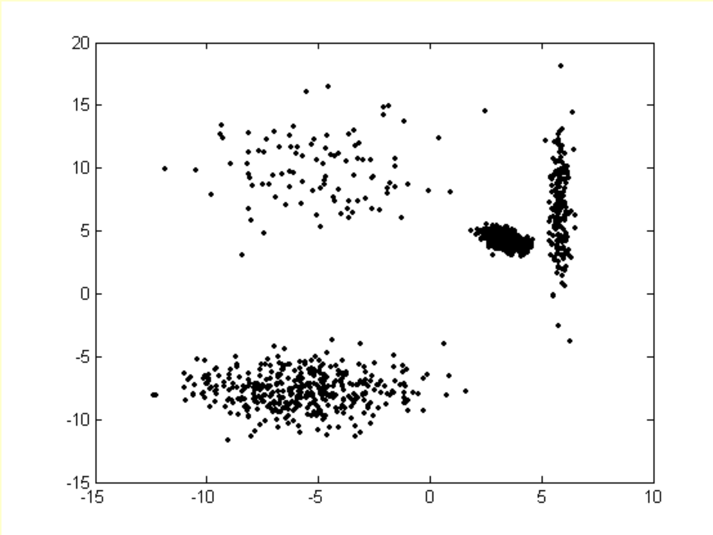



**Hierarchical- single link**

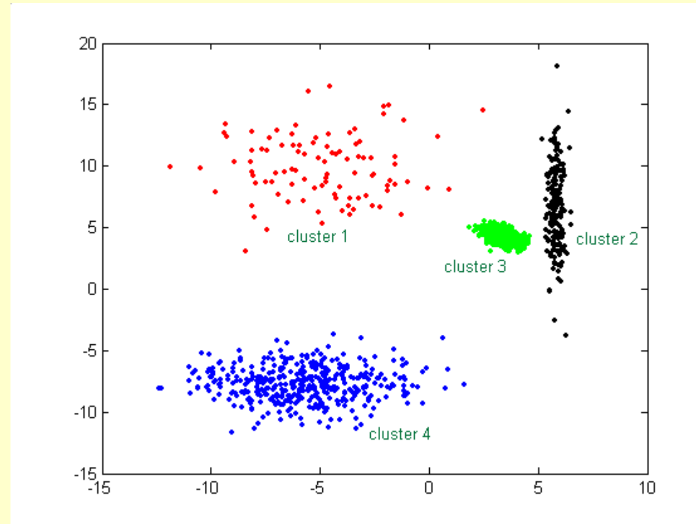



**Hierarchical- complete link**

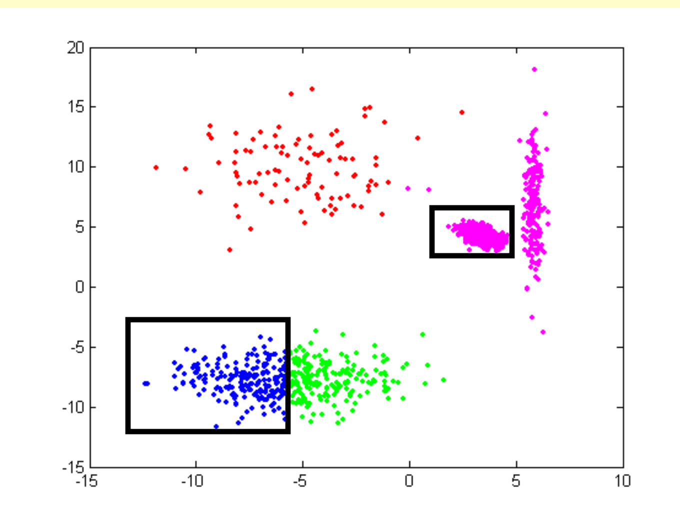
# Clustering results for synthetic data set 4



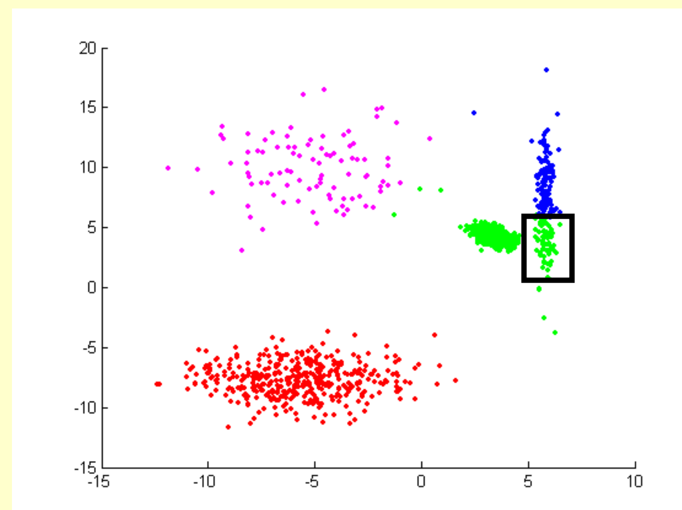
**Un-clustered**




**Ant inspired clustering**

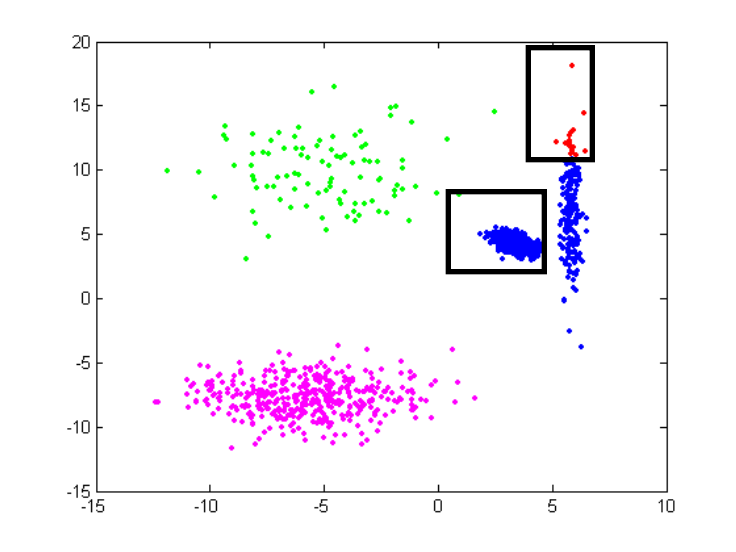


**K means**

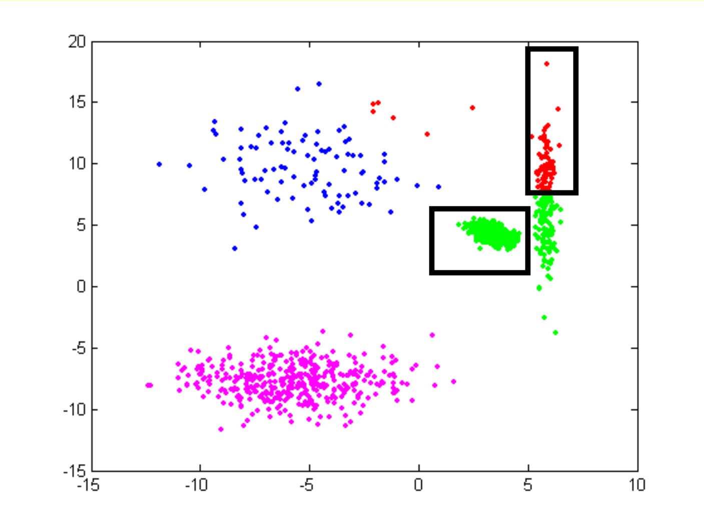


**FCM**

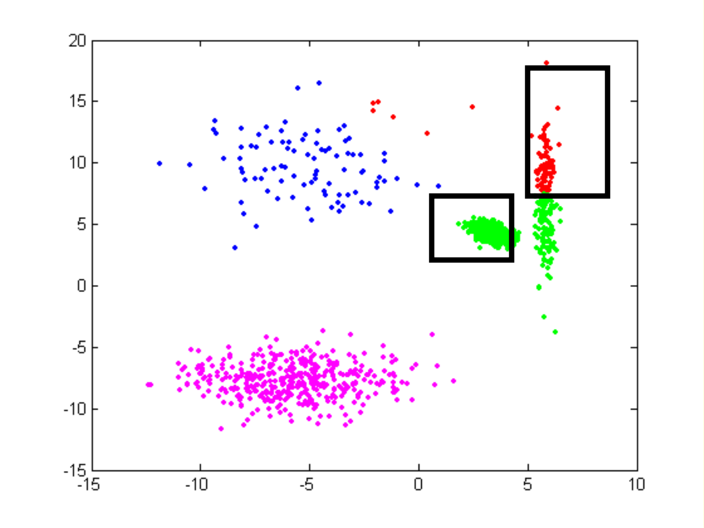
# Clustering results for synthetic data set 4(contd…)




**Hierarchical- average link**




**Hierarchical- single link**




**Hierarchical- complete link**

# Complexity analysis

- Let, $n_i$ be the number of data points left un-clustered to the i-th ant. And if $c_i$ be the no. of data members in the i-th cluster, then the complexity may be given as:-

$$complexity = \sum_i \sum_j \{(3 + j) n_i\}$$

choosing same no. of members for each of the $p$ clusters into which the data may be divided

i.e. $n_i = n - c * (i - 1)$

Then, $complexity = O(n^2 c)$

# Conclusions

1) We presented a very simple algorithm for the very complex clustering problems.

2) The algorithm can detect the correct number of clusters from a virgin dataset "on the run".

3) The algorithm has a quadratic average time complexity.

4) The algorithm can encompass both shell and solid clusters of any arbitrary shapes.

5) However, future research should focus on the adaptation of parameters of the algorithm – the optimal parameter set should be learnt by catching some special statistical features of the data itself.

# References

1) A. K. Jain, M. N. Murty, P. J. Flynn, Data clustering: a review, ACM Computing Surveys, vol. 31, no.3, (1999) 264—323.
2) T Lillesand, R Keifer, Remote Sensing and Image Interpretation, John Wiley & Sons, 1994.
3) H. C Andrews, Introduction to Mathematical Techniques in Pattern Recognition, John Wiley & Sons, New York, 1972.
4) M.R Rao, Cluster Analysis and Mathematical Programming,. Journal of the American Statistical Association, Vol. 22, pp 622-626, 1971.
5) R. O. Duda , P. E. Hart. Pattern Classification and Scene Analysis. John Wiley and Sons, 1973.
6) B.S. Everitt, Cluster Analysis. Halsted Press, third edition (1993).
7) E.W. Forgy, Cluster Analysis of Multivariate Data: Efficiency versus Interpretability of classification, Biometrics, 21, (1965) 768-9.

8) C. T. Zahn, Graph-theoretical methods for detecting and describing gestalt clusters, IEEE Transactions on Computers C-20 (1971), 68–86.
9) T. Mitchell, Machine Learning. McGraw-Hill, Inc., New York, NY (1997).
10) J. Mao, A. K. Jain, Artificial neural networks for feature extraction and multivariate data projection. IEEE Trans. Neural Networks.vol. 6, (1995) 296–317.
11) N. R. Pal, J. C. Bezdek, E. C.-K. Tsao, Generalized clustering networks and Kohonen's self-organizing scheme. IEEE Trans. Neural Networks, vol 4, (1993) 549–557.
12) T Kohonen, Self-Organizing Maps, Springer Series in Information Sciences, Vol 30, Springer-Verlag, (1995).
13) E. Falkenauer, Genetic Algorithms and Grouping Problems, John Wiley and Son, Chichester (1998).
14) S. Paterlini, T. Minerva, Evolutionary Approaches for Cluster Analysis. In A. Bonarini, F. Masulli, G. Pasi (eds.) Soft Computing Applications. Springer-Verlag, Berlin. (2003)167-178.

15) A. Abraham, S. Das, and S. Roy, Swarm Intelligence Algorithms for Data Clustering, *Soft Computing for Knowledge Discovery and Data Mining*, O. Maimon and L. Rokach (Eds.), Springer Verlag, Germany, ISBN 978-0-387-69934-9, pp. 279-313, 2007.
16) J. MacQueen, Some methods for classification and analysis of multivariate observations, Proceedings of the Fifth Berkely Symposium on Mathematical Statistics and Probability, 281-297 (1967).
17) J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, New York, Plenum (1981).
18) S. Bandyopadhyay and U. Maulik, Genetic clustering for automatic evolution of clusters and application to image classification, Pattern Recognition, 35, 1197-1208, 2002.
19) M. Dorigo, V. Maniezzo, and A. Colorni, Ant System: optimization by a colony of cooperating agents, IEEE Transactions on Systems, Man, and Cybernetics–Part B, 26 (1): 29–41, 1996.
20) M. Dorigo and L. M. Gambardella, Ant Colony System: A cooperative learning approach to the traveling salesman problem, IEEE Transactions on Evolutionary Computation, 1 (1): 53–66, 1997.

21) J. Handl and J. Knowles, and M. Dorigo, Ant-based clustering: a comparative study of its relative performance with respect to k-means, average link and 1D-SOM. Technical Report TR/IRIDIA/2003-24. IRIDIA, Universite Libre de Bruxelles, Belgium, 2003.
22) E. Lumer and B. Faieta, Diversity and Adaptation in Populations of Clustering Ants. In Proceedings Third International Conference on Simulation of Adaptive Behavior: from animals to animate 3, Cambridge, Massachusetts, MIT press, pp. 499-508, (1994).
23) M. Omran, A. Salman, and A. Engelbrecht. Dynamic Clustering using Particle Swarm Optimization with Application in Unsupervised Image Classification. Fifth World Enformatika Conference (ICCI 2005), Prague, Czech Republic, 2005.

# DO YOU HAVE ANY QUERIES?

# THANK YOU