Development of a Genetic Algorithm to Automate Clustering of a Dependency Structure Matrix

James L. Rogers
Swales Aerospace, Hampton, Virginia

John J. Korte and Vincent J. Bilardo, Jr.
Langley Research Center, Hampton, Virginia

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA Scientific and Technical Information (STI) Program Office plays a key part in helping NASA maintain this important role.

The NASA STI Program Office is operated by Langley Research Center, the lead center for NASA's scientific and technical information. The NASA STI Program Office provides access to the NASA STI Database, the largest collection of aeronautical and space science STI in the world. The Program Office is also NASA's institutional mechanism for disseminating the results of its research and development activities. These results are published by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers, but having less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services that complement the STI Program Office's diverse offerings include creating custom thesauri, building customized databases, organizing and publishing research results ... even providing videos.

For more information about the NASA STI Program Office, see the following:

- Access the NASA STI Program Home Page at *http://www.sti.nasa.gov*

- E-mail your question via the Internet to help@sti.nasa.gov

- Fax your question to the NASA STI Help Desk at (301) 621-0134

- Phone the NASA STI Help Desk at (301) 621-0390

- Write to:
  NASA STI Help Desk
  NASA Center for AeroSpace Information
  7121 Standard Drive
  Hanover, MD 21076-1320

Development of a Genetic Algorithm to Automate Clustering of a Dependency Structure Matrix

James L. Rogers
Swales Aerospace, Hampton, Virginia

John J. Korte and Vincent J. Bilardo, Jr.
Langley Research Center, Hampton, Virginia

Abstract

Much technology assessment and organization design data exists in Microsoft Excel® spreadsheets. Tools are needed to put this data into a form that can be used by design managers to make design decisions. One need is to cluster data that is highly coupled. Tools such as the Dependency Structure Matrix (DSM) and a Genetic Algorithm (GA) can be of great benefit. However, no tool currently combines the DSM and a GA to solve the clustering problem. This paper describes a new software tool that interfaces a GA written as an Excel macro with a DSM in spreadsheet format. The results of several test cases are included to demonstrate how well this new tool works.

Introduction

In many complex projects, it is helpful to find subsets that have a minimum coupling or interaction with other subsets. These subsets may be organizations, activities, people, or processes. This grouping of subsets is sometimes called clustering, where individual clusters contain most, if not all, of the interactions within the cluster, and the interactions among clusters are reduced or eliminated. Another method is to cluster tightly coupled processes together. This makes it easier to visualize the influences processes might have on one another. There are many applications for clustering ranging from simple problems like clustering the components for a bicycle to make assembly easier; to medium problems such as cluster organizations that closely interact with each other so they may be placed in close proximity to one another; to the more complex such as clusrering the different aspects of a Mars program to aid in assigning personnel twho should work closely together. All of these activities and their couplings can easily be displayed in an Excel^{TM} spreadsheet. A software tool is needed that will rearrange the row and columns of the spreadhseet to cluster like pieces together.

The purpose of this study is to develop an Excel macro that couples a Genetic Algorithm (GA) to a Dependency Structure Matrix (DSM) to cluster tightly coupled processes around the diagonal of the matrix. The DSM is a tool for displaying processes and their couplings (ref. 1.). For an array dimension greater than 10, the manipulation of the rows and columns to cluster along a diagonal is tedious and the convergence process towards a solution is not obvious (which rows to interchange, etc.). The DSM (also referred to as a Design Structure Matrix or a Dependency Structure Method) is based on graph theory where processes (nodes) are placed along the diagonal and their couplings (arcs) are in the off-diagonal elements (ref. 2). The user determines the information contained in the diagonal and off-diagonal elements based on the problem being solved. The DSM is a very flexible tool and has been applied to a wide variety of projects including component-based DSM for modeling system structure based on component interrelationships, people-based DSM for modeling organization structure based on information flow among people in groups, activity-based DSM for modeling project schedule and activity sequencing based on interactivity information flow, and parameter-based DSM for modeling low level relationships between decisions and parameters (ref. 3). The DSM has also been applied to clustering problems (refs. 4 and 5) and cluster processes along the diagonal of the DSM (ref. 6). Much of the research has been done by MIT for developing management tools.

The use of GA's has been instrumental in achieving good solutions to discrete optimization problems that have not been satisfactorily addressed by other methods (ref.7). GA's can rapidly search a very complex design space. The GA has been applied to a DSM problem to find the optimum sequence for the processes of a complex design project based on time and cost (ref.8). GA's have also been applied to clustering problems (refs. 9 and10) such as clustering of like flowers and clustering of patients with multiple sclerosis who have like symptoms. This project combines the DSM and a GA for clustering into a single powerful tool written as an Excel macro. Although GA's and DSM's have been applied separately to clustering problems such as clustering the components of an automobile climate control system,, the literature search indicated that this is the first time a GA has been coupled with the DSM to solve a clustering problem.

There are numerous Excel spreadsheets containing DSM's with processes that design managers wanted to cluster around the diagonal. Before the development of this tool, the process clustering for these problems was typically done by manual manipulation of the DSM. Now, by applying this new tool and changing a few user input parameters, the design manager can rapidly examine many possible combinations and optimize the clustering of the processes around the diagonal.

The Dependency Structure Matrix

The Dependency Structure Matrix (DSM), originally formulated by Steward, is a tool for displaying the sequence of processes (ref. 1). A sample DSM is shown in figure 1.

Systems Integration & Test Branch	1	0	0	1.5	0	0	1.5	1	0.5	1	0	0
Active Sensor Systems Branch	0	2	1	1	2	0	1	1.5	0	0.5	1.5	0.5
Earth & Space Sci. Sys. Analysis/Adv Concepts Br	0	1	3	0.5	1	2	1	0.5	1	0.5	0.5	2
Mechanical Systems Branch	1.5	1	0.5	4	1	0.5	2	1	1	2	0	0.5
Passive Sensor Systems Branch	0	2	1	1	5	0	1	2	0	0.5	1.5	0.5
Adv. Tools & Methods for Analysis and Integr. Br	0	0	2	0.5	0	6	1	0	1.5	0	0	1.5
Structural & Thermal Systems Branch	1.5	1	1	2	1	1	7	1	1	2	0	1.5
Electronics Systems Branch	1	1.5	0.5	1	2	0	1	8	0.5	1.5	2	0
Advanced Engineering Environments Branch	0.5	0	1	1	0	1.5	1	0.5	9	0.5	1.5	1
Exploration Systems Engineering Branch	1	0.5	0.5	2	0.5	0	2	1.5	0.5	10	1	2
Software Systems Branch	0	1.5	0.5	0	1.5	0	0	2	1.5	1	11	0.5
Explor. & Space Tech. Sys. Anal./Adv. Concepts Br	0	0.5	2	0.5	0.5	1.5	1.5	0	1	2	0.5	12

Figure 1. Dependency Structure Matrix

In this DSM, the organization numbers are located the yellow diagonal cells and their couplings are in the off diagonal cells of the matrix. The numbers in the off diagonal cells indicate the strength of the coupling between two organizations. The larger the number, the stronger the coupling is. The strongest couplings are seen in the red off diagonal elements. If the number in the off diagonal cell is 0 then there is no coupling between the two organizations.

Sequence Optimization

Several software tools have been developed to analyze the DSM. One of these tools is called DeMAID (Design Manager's Aid for Intelligent Decomposition, ref.11). The most recent version of DeMAID is called DeMAID/GA to reflect the addition of a Genetic Algorithm (GA) to optimize the ordering of the processes within the iterative subcycles. The fitness function of the GA is based minimizing the time and/or cost of each iterative subcycle. The fitness function uses the time and cost of each process and a factor dependent on the coupling magnitude of each feedback coupling. The resulting DSM will display the processes in the most optimum sequence (top left to bottom right) for execution.

Clustering

In a clustering problem, the DSM is typically symmetric and contains tightly coupled processes. Processes "a" and "b" are considered to be tightly coupled when there are off diagonal elements coupling process "a" to process "b" and process "b" to process "a". The strength of the coupling is indicated by the number in the off diagonal element coupling the two processes, the larger the number the stronger the coupling. For example, in the Langley line-organization study, the coupling was rated a 0, 1, or 2 where 2 indicates a stronger coupling than 1 or 0. The object of the clustering algorithm is not to minimize feedback couplings or to optimize the execution sequence, but optimize the sequence by grouping the tightly coupled processes into clusters around the diagonal. The results can be used to determine which line-organizations should be located near each other.

The Genetic Algorithm

The use of GA's has been instrumental in achieving good solutions to discrete optimization problems that have not been satisfactorily addressed by other methods (ref. 7). The GA searches a population of design points, coded as finite-length, finite-alphabet strings. Successive populations are produced primarily by the operations of selection, crossover, and mutation. Frequently, a binary coding is used with the GA; the values of the design variables are coded as binary numbers (1's and 0's) and then concatenated into a string. While this approach works well with numerical problems, it is not efficient for the clustering problem (refs. 12 and13). This GA uses a direct representation of the order as a coding of an n-process system, with each integer 1 through n used only once.

Selection

The selection operator determines those members of the population that survive to participate in the production of members of the next population. Selection is based on the value of the fitness function, or the fitness of the individual members, such that members with greater fitness levels tend to survive. The tournament selection operator is applied to select members of the mating pool. To fill the mating pool, two strings are randomly selected without replacement from the parent pool and compared (a tournament); the one with greater fitness is included in the mating pool. The same member can be selected more than once and this selection process does not guarantee that the most fit member will be passed along to the mating pool.

Crossover

The crossover operator is the recombination of traits of the selected members, called the mating pool, in the hope of producing a child with better fitness levels than its parents. Crossover is accomplished by swapping parts of the string into which these design points have been coded. Because the strings consist of integers and not just binary 1's and 0's, crossover is accomplished by position-based (ref.11) crossover as shown in Figure 2. Several processes (i.e. 1, 4, 5, and 6) are chosen from the first parent and placed in the same positions in the child string. Then, the processes (i.e. 2, 3, and 7) that were not taken from the first parent are taken from the second parent to fill the holes in the child string in the order in which they appear in the second parent. The result is a complete string with one and only one copy of each process number.

Parent 1 => 1 7 2 4 3 5 6   Parent 2 => 6 5 2 1 3 7 4
Randomly select process numbers 1, 4, 6, and 7 from Parent 1

Child => 1 _ _ 4 _ 5 6
Fill in from Parent 2 based on encountering missing process numbers
Child => 1 2 3 4 7 5 6

Figure 2. Position Based Crossover Function

Mutation

The mutation operator prevents the search of the space from becoming too narrow or getting hung up in a local minimum. After the production of a child population, this operator randomizes small parts of the resulting strings, with a very low probability that any given string position will be affected. Mutation is accomplished through the order-based (ref. 11) mutation operator, as shown in Figure 3. Each string position is polled; if a given string position (i.e. position 2) is randomly selected to undergo mutation, then its content is swapped with a randomly selected position (i.e. position 4) in the same string

String before mutation => 1 7 2 4 3 5 6
String after mutation => 1 4 2 7 3 5 6

Figure 3. Order Based Mutation

Fitness Function

The previous GA operators, selection, crossover, and mutation are somewhat problem independent. The fitness function, however, is typically problem dependent. For the GA used for clustering, the GA attempts to move all of the highest valued off diagonal elements of the DSM as close to the diagonal as possible. For example, in the DSM in figure 1, the user defined a cluster parameter of 1.5, so no off diagonal element with a value less than 1.5 will be considered in computing the fitness function. The off diagonal elements that will be part of the calculation of the fitness function are in red. The user also defines the value for penalizing the fitness function, for example 10. The fitness function routine loops through all cells in the matrix and for each off diagonal cell that has a value greater than the cluster parameter (1.5 in figure 1) a fitness value is calculated. This calculation multiplies the value of the off diagonal element by the absolute value of the difference between the column of the element and the diagonal. Then this product is multiplied by the penalty value so that the farther away an element is from the diagonal then the larger the fitness function. For example, suppose that an off diagonal element with a value of 2 is in row 2 column 10 and the cluster penalty is 10. The fitness for that element would be 2*(abs(10-2)*10 or 160. However, if the GA moved that element to row 4, then the fitness would be reduced to 2*abs(4-2)*10 or 40. All of these penalties are summed to create the fitness value. Thus the GA tends to move the highest valued off diagonal elements closer to the diagonal and that moves the highly coupled processes closer to one another in the sequence. The graph in figure 4, demonstrates how the fitness function changes over the generations. Sometimes, there is a significant reduction, while other times there are no changes for several generations, followed by a slight reduction. This graph is from the Langley Organization test case.
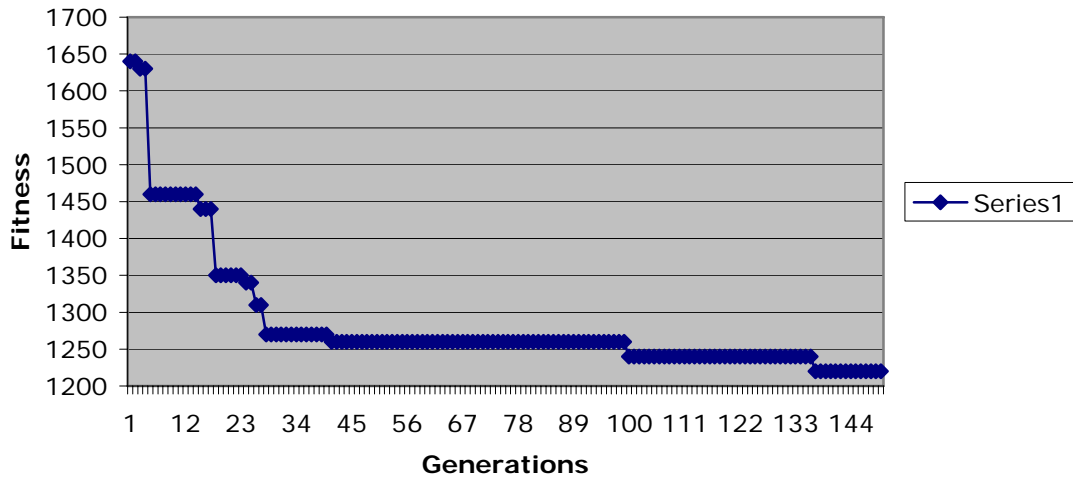
Figure 4. Fitness Function History

Applying the Genetic Clustering Algorithm

This clustering GA is written in Microsoft Visual Basic® as an Excel macro. This enhances portability as well as facilitates interaction with existing DSM's in Excel spreadsheet format. There are several aspects to consider when applying this GA. Every aspect is accomplished within an Excel Spreadsheet.

Data input

The user inputs the problem data in certain cells in column 1 of the Original DSM worksheet as shown in Figure 5.

Input Data
Maximum Generations
50
Population Size
100
Maximum Processes
10
Mutation Rate
0.01
Cluster Parameter
3
Cluster Penalty
10
Trace Fitness Function

Figure 5. Worksheet Input Values

The cell in row 1 column 1 is not to be used because it stores a random number for the random number generator. The labels for each of the parameters are in the row immediately above the

parameter. The user defines certain GA parameters such as the maximum number of generations, not to exceed 500 in row 5 (50), the population size, not to exceed 500 (a rule of thumb is to make the population size about 10 times the number of maximum processes but do not exceed the maximum number for population size) in row 8 (100), and the mutation parameter in row 14 (1% = .01). The user defines the maximum number of processes, not to exceed 200 in row 11 (10), and fitness function values for the cluster parameter (the lowest number to be included in computing the fitness of a string) in row 17 (3) and the cluster penalty indicating how much an element is penalized by being away from the diagonal, in row 20 (10). The numbers in parentheses above reflect the numbers in Figure 5 for each parameter. The GA keeps track of the fitness value as it proceeds through each generation. These values are shown in column 1 beginning with row 23. Before executing the genetic algorithm, the user is to store an existing DSM beginning with the process names in row 1, column 2, and the data beginning in row 1, column 3.

Macro Execution

When opening the ClusterGA Excel file, the user will be asked to Enable Macros. Click the button to enable them. To execute the cluster GA, the user selects Macro from the Tools menu, and then selects Macros from the Macro menu. Highlight the ClusterGA macro name and the click the Run button. The resulting DSM will appear on the Cluster DSM worksheet. Off diagonal cells used for clustering will be red on both the Original and Clustered DSM worksheets. In addition, to the right of the Clustered DSM, there will be a column indicating the position of that process in the Original DSM.

Manual Manipulation of DSM Processes

If the user wishes to move some processes around the DSM manually, the user can return to the tools menu and again select Macro and Macros. This time, highlight the MoveProcesses macro and click the Run button. The user will be asked which process to move and which process to move it after. The macro will then adjust the rows and columns to make the change. This macro continues until the user selects no more processes to be moved. All manual changes are stored on the Final DSM worksheet. No changes are made to the other two worksheets.

Test Cases

The clustering GA was tried on several test cases. The effect of clustering with the GA is demonstrated by showing before and after example cases. The yellow cells (or light gray) are on the diagonal. The red cells (or dark gray) indicate all off diagonal cells containing a value used in determining the fitness function. Each test case was executed with these parameters: population size =300, maximum generations = 400, mutation rate = .01, and cluster penalty = 10. The number of processes and the cluster parameter varied with each test case. Each DSM is reordered by the GA indicate stronger coupling of the off diagonal elements under consideration around the diagonal. (Note: The spreadsheets in Figures 8, 12 and 14 are from Brady, ref. 6)

Test Case 1- A DSM was constructed to evaluate the degree of communication between different branches from two different organizations at the NASA Langley Research Center. Figure 6 shows the original DSM filled in with the original scores. There were 12 processes and the cluster parameter was 1.5. The clustering GA was applied to the DSM in Figure 6 and the results are shown in Figure 7. This clearly shows that the GA provided the desired behavior for clustering along the diagonal.

Systems Integration & Test Branch

1	0	0	1.5	0	0	1.5	1	0.5	1	0	0
0	2	1	1	2	0	1	1.5	0	0.5	1.5	0.5
0	1	3	0.5	1	2	1	0.5	1	0.5	0.5	2
1.5	1	0.5	4	1	0.5	2	1	1	2	0	0.5
0	2	1	1	5	0	1	2	0	0.5	1.5	0.5
0	0	2	0.5	0	6	1	0	1.5	0	0	1.5
1.5	1	1	2	1	1	7	1	1	2	0	1.5
1	1.5	0.5	1	2	0	1	8	0.5	1.5	2	0
0.5	0	1	1	0	1.5	1	0.5	9	0.5	1.5	1
1	0.5	0.5	2	0.5	0	2	1.5	0.5	10	1	2
0	1.5	0.5	0	1.5	0	0	2	1.5	1	11	0.5
0	0.5	2	0.5	0.5	1.5	1.5	0	1	2	0.5	12

Row labels (top to bottom): Systems Integration & Test Branch; Active Sensor Systems Branch; Earth & Space Sci. Sys. Analysis/Adv Concepts Br; Mechanical Systems Branch; Passive Sensor Systems Branch; Adv. Tools & Methods for Analysis and Integr. Br; Structural & Thermal Systems Branch; Electronics Systems Branch; Advanced Engineering Environments Branch; Exploration Systems Engineering Branch; Software Systems Branch; Explor. & Space Tech. Sys. Anal./Adv. Concepts Br

Figure 6. Langley Organization Test Case (before GA)

1	2	2	0.5	1	0.5	0	1	0.5	0.5	1	1
2	2	1.5	0	1	0.5	0	1.5	0	0	0	0
2	1.5	3	2	1.5	0.5	0	1	0.5	0	0.5	0.5
0.5	0	2	4	2	2	1	0.5	1	1.5	0.5	0.5
1	1	1.5	2	5	2	1.5	1	0	1	1	1
0.5	0.5	0.5	2	2	6	1.5	1	0	1	1	1
0	0	0	1	1.5	1.5	7	0.5	0	1	0	0
1	1.5	1	0.5	1	1	0.5	8	1.5	0.5	0	0
0.5	0	0.5	1	0	0	0	1.5	9	2	1.5	1.5
0.5	0	0	1.5	1	1	1	0.5	2	10	2	1.5
1	0	0.5	0.5	1	1	0	0	1.5	2	11	2
1	0	0.5	0.5	1	1	0	0	1.5	1.5	2	12

Row labels (top to bottom): Earth & Space Sci. Sys. Analysis/Adv Concepts Br; Adv. Tools & Methods for Analysis and Integr. Br; Explor. & Space Tech. Sys. Anal./Adv. Concepts Br; Exploration Systems Engineering Branch; Structural & Thermal Systems Branch; Mechanical Systems Branch; Systems Integration & Test Branch; Advanced Engineering Environments Branch; Software Systems Branch; Electronics Systems Branch; Passive Sensor Systems Branch; Active Sensor Systems Branch

Figure 7. Langley Organization Test Case (after GA)

Test Case 2-This example was taken from Brady (ref. 6). It demonstrates the grouping of bicycle components into subsystems. The original DSM is given in Figure 8 and represents the strength of the interfaces among the different components. There were 10 processes and the cluster parameter was 4. Figure 9 shows the results of applying the GA to the DSM. The Excel spreadsheet indicates the original numbers for each process in a separate column.

	Wheels	Gears	Chain	Pedals	Handlebars	Frame	Brake	Brake Handle	Gear Shift	Odometer
Wheels	1	4				2	4			2
Gears	4	2	4						4	
Chain		4	3	4						
Pedals			4	4		2				
Handlebars					5	2		2	2	2
Frame	2			2	2	6	2	2	2	
Brake	4					2	7	4		
Brake Handle					2	2	4	8		
Gear Shift		4			2	2			9	
Odometer	2					2				10

Figure 8. Bicycle test case (before GA)

	Pedals	Chain	Gear Shift	Gears	Wheels	Brake	Brake Handle	Handlebars	Frame	Odometer
Pedals	1	4							2	
Chain	4	2		4						
Gear Shift			3	4				2	2	
Gears		4	4	4	4					
Wheels				4	5	4			2	2
Brake				4	6	4		2		
Brake Handle						4	7	2	2	
Handlebars			2				2	8	2	2
Frame	2		2		2	2	2	2	9	
Odometer					2			2		10

Figure 9. Bicycle test case (after GA)

Test Case 3-This example was taken from the Next Generation Launch Vehicle (NGLV) study. The original DSM is given in Figure 10 and represents the strength of the interfaces for the different organizations composing the NGLV task. There were 46 processes and the cluster parameter was 6. Due to size restrictions, these numbers cannot be shown in the figure. Figure 11 shows the results of applying the GA to the DSM.
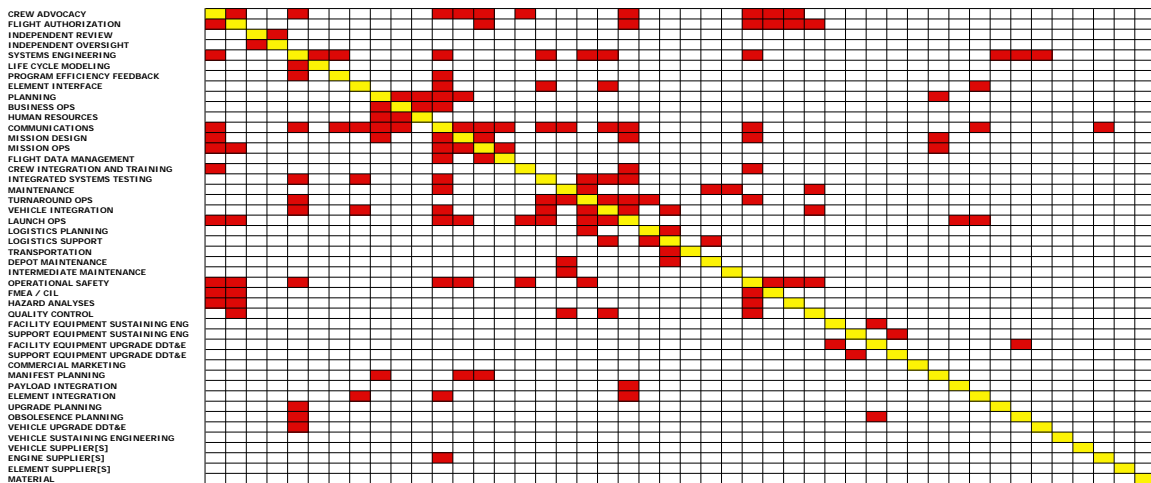
Figure 10. NGLV Organization (before GA)
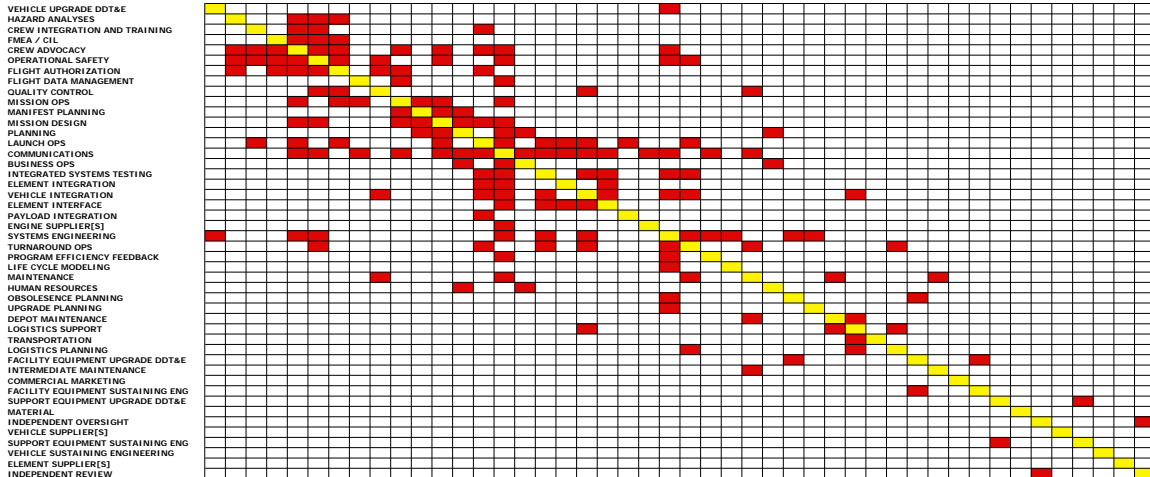
Figure 11. NGLV Organization (after GA)

Test Case 4-This example was taken from Brady (ref. 6). It demonstrates the dependency of the components for Mars Global Surveyor associated with propulsion, attitude control, telecommunications, and science instruments. The original DSM is given in Figure 12 and represents the interfaces among the different components based on a risk factor. There were 26 processes and the cluster parameter was 4. Due to size restrictions, these numbers cannot be shown in the figure. Figure 13 shows the results of applying the GA to the DSM.
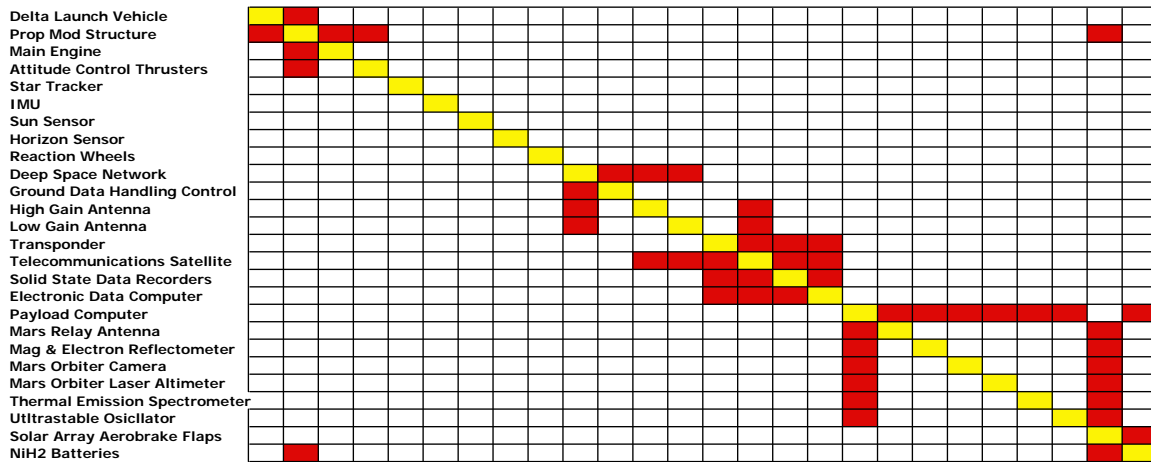
Figure 12. Mars Global Surveyor (before GA)
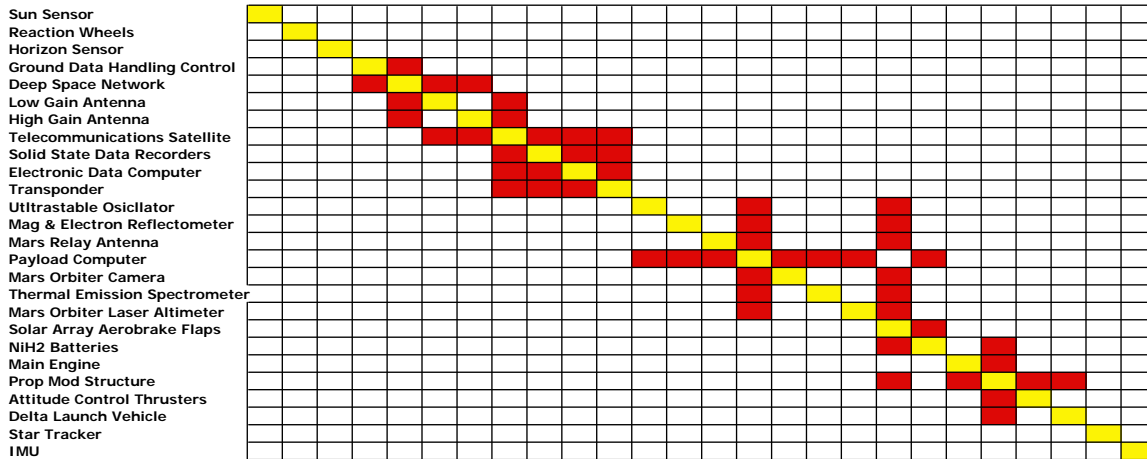
Figure 13. Mars Global Surveyor (after GA)

Test Case 5-This example was taken from Brady (ref. 6). It demonstrates the components of the Mars Global Surveyor associated with propulsion, telecommunications, power and payload. The original DSM is given in Figure 14 and represents the represents the interfaces among the different components based on a risk factor. There were 20 processes and the cluster parameter was 4. Due to size restrictions, these numbers cannot be shown in the figure. Figure 15 shows the results of applying the GA to the DSM.
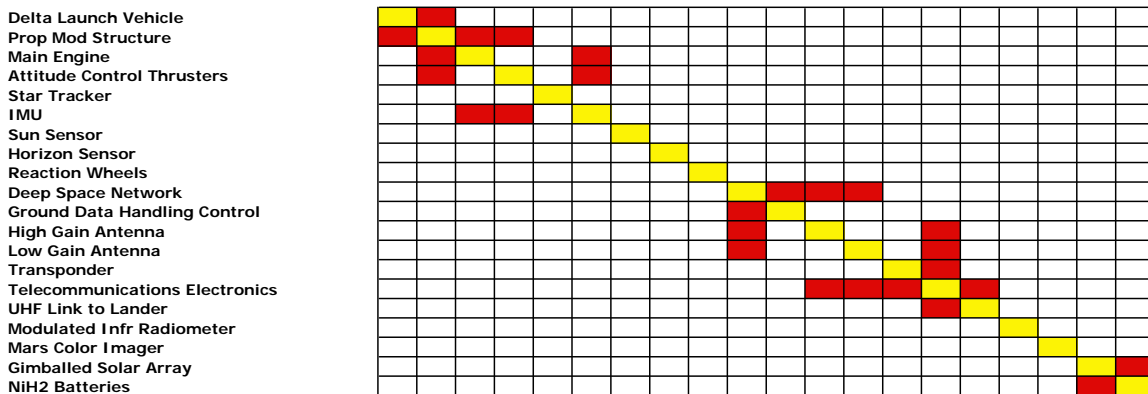
Figure 14. Mars Climate Orbiter (before GA)
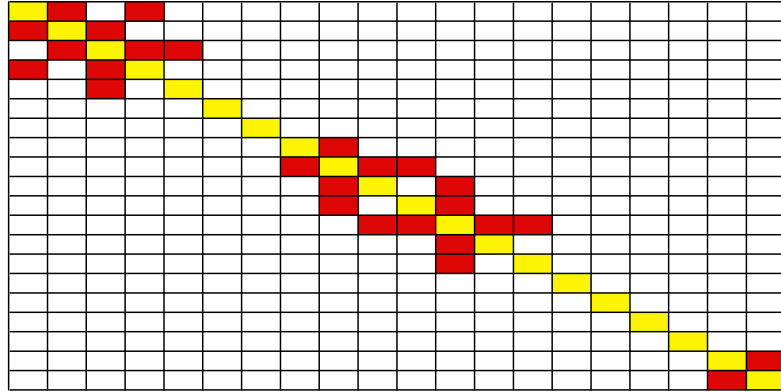
Figure 15. Mars Climate Orbiter (after GA)

Test Case 6-This example was taken from data generated by a Skills Assessment Team... The Skills Assessment Team filled out a DSM focusing on in-house skills needed to provide technical and management insight leading to a down select for the Crew Exploration Vehicle (CEV). The original DSM is given in Figure 16 and represents the strength of the interfaces among the different skills being assessed. There were 36 processes and the cluster parameter was 7. Due to size restrictions, these numbers cannot be shown in the figure. Figure 17 shows the results of applying the GA to the DSM.
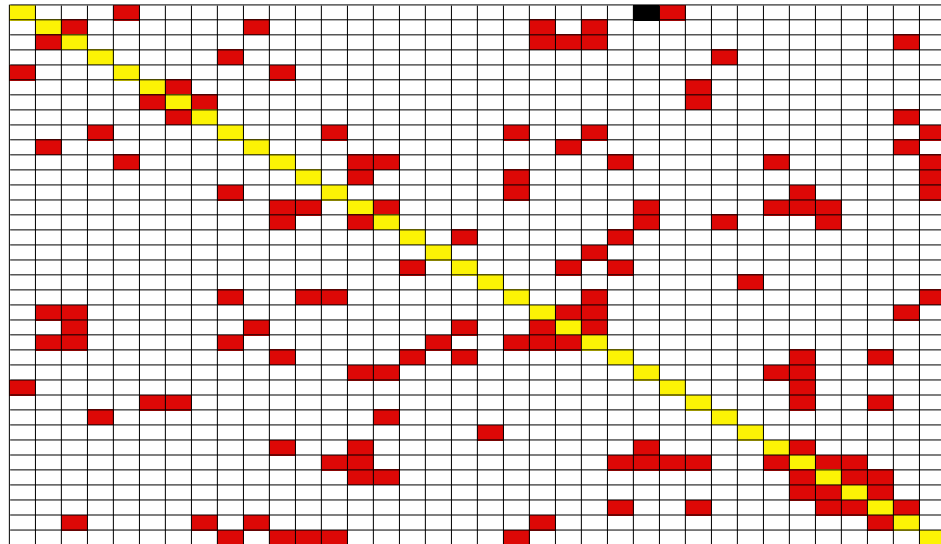
Figure 16. Skills Assessment (before GA)

11
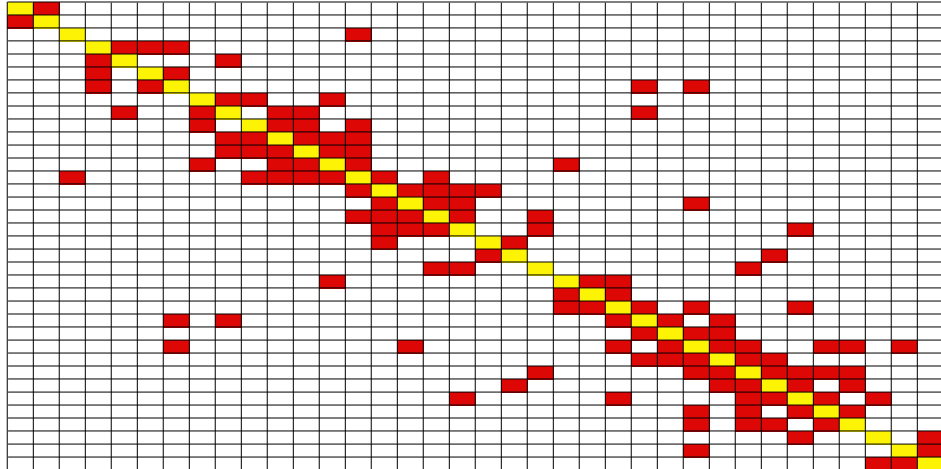
Figure 17. Skills Assessment (after GA)

Test Case 7-This example was taken from skills data generated by the Investigation Definition Team (IDT)... The Investigation Definition Team filled out a DSM focusing on in-house skills needed to provide technical and management insight leading to a down select for the Crew Exploration vehicle (CEV). The original DSM is given in Figure 18 and represents the strength of the interfaces among the different skills being assessed. There were 64 processes and the cluster parameter was 3.9. Figure 19 shows the results of applying the GA to the DSM.
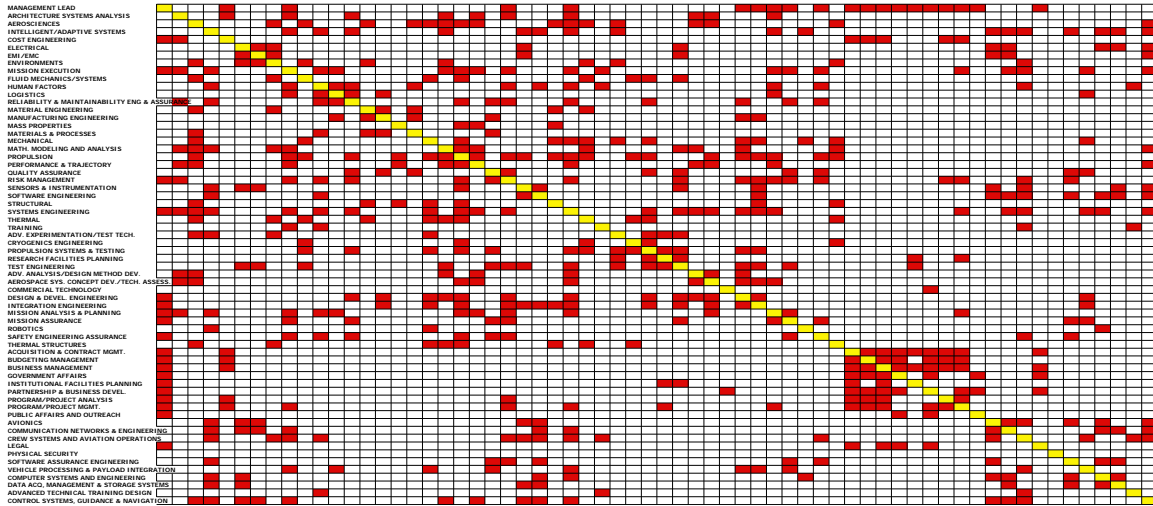
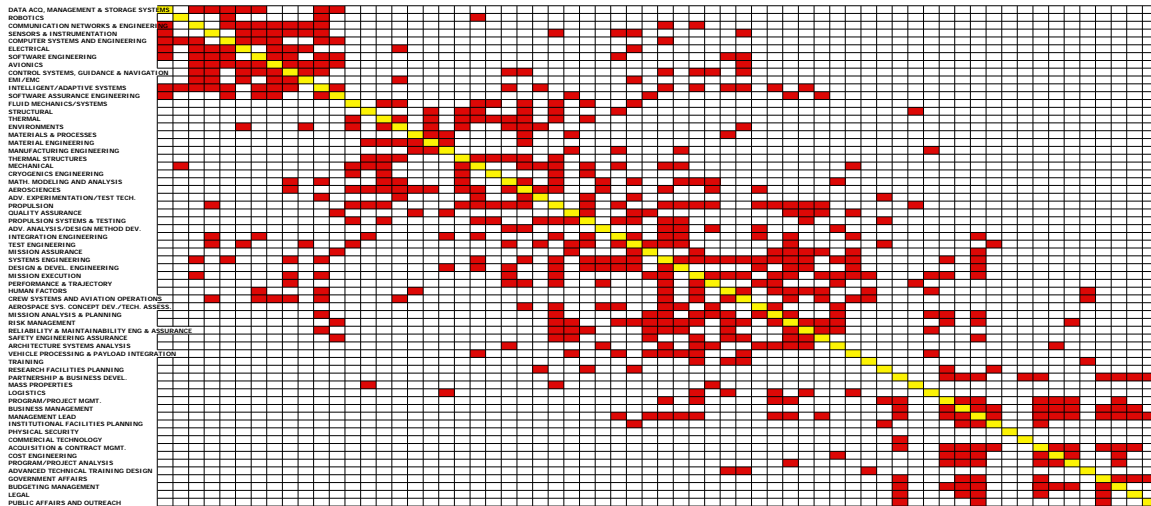

Figure 18. IDT Formulation (before GA)

Figure 19. IDT Formulation (after GA)

Summary

Numerous DSM's exist in Excel spreadsheet format. Some of them contain data that can be used in technology assessment. Design managers needed to cluster the tightly coupled processes along the diagonal to better visualize the impact they have on one another. Other than using the tedious method of moving the processes manually, no tool existed to aid in solving this clustering problem.

A literature search found that GA's had been applied to clustering problems and in optimizing the sequence of processes displayed in DSM format. It was also found that DSM's had been applied to clustering problems. However, no tools were found that coupled the GA and the DSM to solve the clustering problem.

A GA was coded in Visual Basic as an Excel macro to interface with existing DSM's in Excel spreadsheet format. The GA was tested with a variety of DSM's. Some of the DSM's were small (10 processes) and some were much larger (64 processes). Some of the DSM's were sparsely populated with data while others were densely populated. In each case, the resulting optimized DSM had moved the tightly coupled processes to be clustered around the diagonal.

References

1. Steward, D. V.: Systems Analysis and Management: Structure, Strategy and Design. Petrocelli Books Inc., New York. 1981.

2. Rogers, James L.: Tools and Techniques for Decomposing and Managing Complex Design projects. Journal of Aircraft, Vol. 36 No. 1, Jan.-Feb. 1999, pp. 266-274.

3. Browning, T. R.: Dependency Structure Matrix-Based Systems Engineering, Organization Planning, and Schedule Management, M.I.T. Sloan School of Management, Cambridge, MA, Working Paper March 1995, revised February 1998.

4. Fernandez, C.I.G.: Integration Analysis of Product Architecture to Support Effective Team Co-Location, Master's Thesis (ME), Massachusetts Institute of Technology, Cambridge, MA, 1998.

5. Thebeau, R. E.: Knowledge Management of System Interfaces and Interactions for Product Development Processes, MIT Masters Thesis, System Design & Management Program, Feb. 2001.

6. Brady, T. K.: Utilization of Dependency Structure Matrix Analysis to Assess Implementation of NASA's Complex Technical Projects, MIT Master's Thesis, February 2002.

7. Goldberg, D.: Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley Publishing Co., New York, 1989.

8. Rogers, J. L.; McCulley, C. M.; and Bloebaum, C. L.: Integrating a Genetic Algorithm into a Knowledge-Based System for Ordering Complex Design Processes, Proceedings of the Artificial Intelligence in Design '96 Conference, Stanford, CA, June 1996, pp. 119-133. (Also NASA TM 110247 April 1996)

9. Bezdek, J. C. and Hathaway, R. J.: Optimization of Fuzzy Clustering Criteria using Genetic Algorithms, Proc. First IEEE Conference on Evolutionary Computation, Vol 2, pp. 589-594, 1994.

10. Bhuyan, J. N., Raghavan V. V., and Elayavalli, V. K.: Genetic Algorithm for Clustering with an Ordered Representation, Proc. International Conference on Genetic Algorithms '91, pp. 408-415, 1991.

11. Rogers, James L.: DeMAID/GA User's Guide---Design Manager's Aid for Intelligent Decomposition With a Genetic Algorithm, NASA TM-110241, April 1996.

12. Altus, S. S.; Kroo, I. M.; and Gage, P. J.: "A Genetic Algorithm for Scheduling and Decomposition of Multidisciplinary Design Problems", ASME Paper 95-141. 1995.

13. McCulley, C. M.; and Bloebaum, C. L.: "Optimal Sequencing for Complex Engineering Systems Using Genetic Algorithms." Fifth AIAA/USAF/NASA/OAI Symposium on Multidisciplinary Analysis and Optimization, Panama City, FL. 1994.

REPORT DOCUMENTATION PAGE

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
01- 02 - 2006	Technical Memorandum	

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Development of a Genetic Algorithm to Automate Clustering of a Dependency Structure Matrix	
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)	5d. PROJECT NUMBER
Rogers, James L.; Korte, John J.; and Bilardo, Vincent J., Jr.	
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
	23-617-99-01-TL

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
NASA Langley Research Center Hampton, VA 23681-2199	L-19157

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
National Aeronautics and Space Administration Washington, DC 20546-0001	NASA
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)
	NASA/TM-2006-214279

12. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified - Unlimited
Subject Category 59
Availability: NASA CASI (301) 621-0390

13. SUPPLEMENTARY NOTES
An electronic version can be found at http://ntrs.nasa.gov

14. ABSTRACT

Much technology assessment and organization design data exists in Microsoft Excel® spreadsheets. Tools are needed to put this data into a form that can be used by design managers to make design decisions. One need is to cluster data that is highly coupled. Tools such as the Dependency Structure Matrix (DSM) and a Genetic Algorithm (GA) can be of great benefit. However, no tool currently combines the DSM and a GA to solve the clustering problem. This paper describes a new software tool that interfaces a GA written as an Excel macro with a DSM in spreadsheet format. The results of several test cases are included to demonstrate how well this new tool works.

15. SUBJECT TERMS
DSM; Dependency Structure Matrix; Genetic Algorithm; Organizational Analysis

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			STI Help Desk (email: help@sti.nasa.gov)
					19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	19	(301) 621-0390