SANDIA REPORT

SAND2003-8754
Unlimited Release
Printed January 2004

Computational Social Dynamic Modeling of Group Recruitment

N. Berry, T. Ko, M. Lee, T. Moy, M. Pickett, J. Smrcka, J. Turnley, B. Wu

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94-AL85000

Approved for public release; further dissemination unlimited.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865)576-8401
Facsimile: (865)576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone: (800)553-6847
Facsimile: (703)605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2003-8754
Unlimited Release
Printed January 2004

Computational Social Dynamic Modeling of Group Recruitment

Nina Berry, Teresa Ko, Marinna Lee, Marc Pickett, Ben Wu
Sandia National Laboratories
P.O. Box 969
Livermore, CA 94551-0969

Timothy Moy, Julianne Smrcka, Jessica Turnley
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-9012

Abstract

The Seldon software toolkit combines concepts from agent-based modeling and social science to create a computationally social dynamic model for group recruitment. The underlying recruitment model is based on a unique three-level hybrid agent-based architecture that contains simple agents (level one), abstract agents (level two), and cognitive agents (level three). This uniqueness of this architecture begins with abstract agents that permit the model to include social concepts (gang) or institutional concepts (school) into a typical software simulation environment. The future addition of cognitive agents to the recruitment model will provide a unique entity that does not exist in any agent-based modeling toolkits to date. We use social networks to provide an integrated mesh within and between the different levels. This Java based toolkit is used to analyze different social concepts based on initialization input from the user. The input alters a set of parameters used to influence the values associated with the simple agents, abstract agents, and the interactions (simple agent-simple agent or simple agent-abstract agent) between these entities. The results of phase-1 Seldon toolkit provide insight into how certain social concepts apply to different scenario development for inner city gang recruitment.

This page intentionally left blank

Contents

Figures

Tables

1 Introduction

1.1 Seldon Overview

The tragedy of 9/11 firmly gave the nation and its supporting resources a new collection of challenges associated with the war on terrorism. One strategic necessity in meeting these challenges is a better understanding of the social dynamics behind the violent and aggressions behavior of terrorist and terrorist-like organizations. The pursuit of research in this area covers a collection of multi-disciplinary fields including sociology, psychology, agent-based technology, modeling, simulation, and cognitive science.

For years Sandia has been involved in ongoing research and development in the areas of agent-based modeling for economics, and electrical grid, and system security. Continued interest in this technology has led Sandia to sponsor the agent-based swarm system research at the Santa Fe Institute Complexity System Division. Research developments in cognitive science modeling at Sandia have led to a newly funded Grand Challenge focusing on the issues of cognition and its application to NWIE issues. These efforts have ideally positioned Sandia to capitalize on the intersection of agent-based technology, modeling and cognition with the concepts of social science to enhance the evolving field of computational social dynamics.

Computational social dynamics is a rapidly emerging transdisciplinary field that combines computational and social sciences forming complex adaptive agents that can be applied to simulation models of societal dynamics. Social scientists have long recognized the critical interdependence of individuals on society and vice versa, but up to now there has not been a computational tool for analyzing such non-linear interactions in such complex social systems [19]. This interdisciplinary effort allows for social scientist to have a laboratory to manipulate variables that are not possible in human research. The concept of compression that is inherent in modeling and simulation challenges social scientist. Social scientist thinking is contextual in nature, identifying all environmental influences to a dynamic system. These influences have included socioeconomic status, geography (i.e. urban versus rural),, racial/ethnic, biological, governmental/institutional, as well as tribal affiliations. [18]. The social scientist if forced to synthesize the most critical component and converting behavior into algebraic equations.

The Seldon project (and model) takes its name from Hari Seldon, the fictitious originator of "psycho-history" in Isaac Asimov's Foundation stories. In those stories, Seldon was able to employ a deep knowledge of history, the social sciences, and mathematics to forecast large-scale and long-term trends in the development of civilization. Only large-scale forecasting was possible; in fact, the major tension in the stories stems from the unexpected role of a single unique individual, born long after Seldon's death, who threatens to disrupt Seldon's calculations and destroy the plans he based upon them.

Today, psycho-history remains a science fiction fantasy. However, computational tools are now being created that allows short-term modeling of tightly constrained, complex social interactions. Unlike psycho-history, these tools cannot provide unique insight and understanding. Nevertheless, they often illuminate social science dynamics arising from the interactions of individuals in groups. These insights can help us anticipate classes or types of possible terrorist activities. These tools also can help us understand the broad consequences of our efforts to interdict and mitigate these activities.

The application of computer simulations to social science problems began in the early 1960's and has progressively moved toward finer-grained forms of social simulation in the late 1990's. The earliest computer simulation techniques dealt with macro-simulation, which addressed modeling at a global level with total population distributions in cities and industry. The desire to explore social issues from the bottom-up next led the field into the development of micro-simulation, exploring the interactions between smaller decision making groups within an organization, like families or firms [1]. While both macro and micro simulation provided uniquely different solutions, this progression illustrates the desire to capture changes within the simulation from coarser to finer granularities of social activity, respectively.

The new migration towards integrating agent-based technology and social science concepts will provide a finer granularity of simulation that is based on interactions between actual agents [1]. The ability to observe simulations at this level could provide unique insight into how social relationships, societal rules, and environmental factors are integrated into a complete world model. While these simulations cannot be used to predict the behavior of specific individuals at defined points in time or space, they provide a tool for understanding social behavior in a variety of situations.

The primary benefit of such a tool would be to provide a unique computational gaming environment where one can evaluate the effectiveness of strategic interventions on the emergence and persistence of terrorist groups. In the absence of such computational tools, intrinsically non-linear social systems would be challenging to understand because of both the multiple interdependencies of the processes and the adaptive nature of the individuals. The long-term vision of a tactical terrorist model will require a significant and concerted effort from a transdisciplinary team borrowing from both physical and social sciences.

1.2 Objective

The objective of this LDRD project was to develop a prototype social dynamic toolkit that enabling us to capture a minimal set of social structures, processes, and features analogous to terrorist groups. While the primary target was ultimately terrorist organizations, the lack of terrorist data precluded model development based solely on the domain of terrorist behavior. The domain experts on our team had previously identified domestic urban street gangs as a candidate for this role because of their strong (although not perfect) analogy to many salient dimensions of terrorist groups. We therefore used urban street gangs as a terrorist surrogate in the development of this software tool.

1.3 Deliverables

This LDRD deliverable was a prototype agent-based computational modeling toolkit for simulating the emergence of terrorists and terrorist-like organizations. More specifically, our tasks were to:

- Design and develop domain expert model for terrorist analog (urban street gangs)
- Design and develop architecture for modeling program
- Design and develop preliminary agent-based social simulation toolkit.

2 Background: Computation Social Simulations

The theoretical basis of some types of computational social simulation has been around for decades. Social network analysis, for example, emerged in the 1930's with the combination of certain social and social psychological approaches with graph theory. Agent-based modeling is more recent, but still has a decades-long history. It applies models of the physical and life sciences to social phenomena, seeking particularly to explain phenomena of 'emergence' and complexity.

While these models are all very powerful, it is important to briefly discuss their limitations. Some of these limitations restrict the type of social phenomena that can be richly explored by these models. Others will predispose various potential users to discount their utility.

Most computational social models default to a strong methodological individualism, that is, it is only individuals that are 'real.' Group phenomena are functions of the actions of individuals (they 'emerge') and have no ontological status in their own right. The models are generally postivist in nature and only consider observable phenomena as data. This completely precludes the inclusion of the semiotic dimension (symbols and meaning) that many argue is uniquely human. Finally, the models do not address any degree of autonomy or free will for the agents. Again, these limitations do not negate the models' utility; they simply bound it.

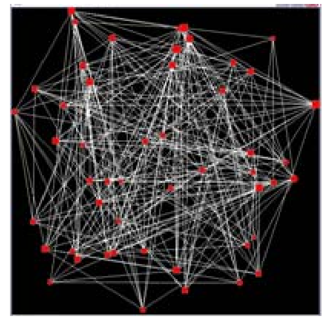
2.1 Background: Agent-based Modeling

The concepts behind earlier agent-based modeling dealt mainly with organizational issues between large and medium organizations or resources. These macro-level simulations covered business interactions between banks, insurance companies, power plants (politics, ancient societies as example of human organization), etc. The next migration dealt with micro-level capabilities of the agent entities and covered internal organizational issues of banks and other companies as needed for the simulation model.

Today these approaches have evolved into fine-grained representations of individuals who interact with each other and their surroundings. The degree of granularity used to represent these individual agents (autonomous, interactive, reactive and proactive) is not an exact science and differs heavily from one implementation to the next. Decisions governing the representation of the agent are usually based on the application domain. The only known limitations is that researchers would not attempt to replicate an "actual" human with the technology and techniques used in current agent-based modeling research.

The newest advancement in agent-based modeling is the incorporation of social networks into the model. In earlier systems used social interactions that were defined by unrelated graph theory instead of real social data and organizational theory. Today researchers are starting to investigate and incorporate social theory and data on relationship structures in these modeling techniques. This new addition, known as social networks, is a major difference between ant-like agent-based modeling and the newest approaches in this field.

2.2 Social Networks



One of the distinguishing features of the Seldon model is that we impose social network(s) that dictate the types and frequencies of agent-agent interactions, virtually absent in other computational social simulations. Typically, agent interactions are purely stochastic in nature, effectively random meetings between two particles. This probabilistic approach neglects the sociological structures that are critical to both constraining and facilitating interactions in human communities. In other words, humans most frequently interact with other humans with whom they already have an established link as part of a network, e.g. - bowling buddies.

There is also the possibility that two unfamiliar agents can interact outside of established networks, for instance a chance meeting at the grocery store. We allow for both types of interactions: familiar and unknown, and we will show that imposing a social network significantly alters the simulation output and should therefore not be neglected. It is important to note that the existence of social networks distinguishes human social dynamics from those of insects that have been frequently cited in popular literature.

In the Seldon model, the teen agents can belong to four different social networks: (1) school attendees/strangers (2) truants/strangers, (3) gangs, and (4) friends. Networks 1 and 2 are fully connected, so that all agents can interact. The interpretation of this computational approach is that these interactions between unfamiliar agents are effectively probabilistic, like a chance meeting at the cafeteria. Network 3 (gangs) is also a fully connected network although for the opposite reason. These networks are sufficiently intimate that each of the agents can interact with all other agents in that network.

Unlike the other three networks, Network 4 ("friend") allows for only limited connectivity between agents, with approximately 2-50 links per agent. The limited network simulates the cliques that, in the absence of gangs, guide most of the daily agent-to-agent interactions. The "friend" network was constructed randomly during the model initialization and remained unchanged for the simulation.

The selection of a randomly generated network for Network 4 was a difficult choice made mostly for lack of a better option after consultations with Professor Kathleen Carley. We briefly describe in the following section the two possible structures for social networks.

Are social networks random, power-law, or neither?

Much has been recently written on scale-free networks [2, 5, 6, and 7] and their apparent applicability to describing wide range of complex networks far more accurately than random networks. The primary differences are illustrated in the figure below taken from Barabási et al. [7]. The top two illustrations (c) and (e) are typical of a random network where *P(k)* is the probability that a randomly selected node has *k* edges. The normal distribution is characteristic of a random network. In contrast, the scale-free network of (d) and (e) is characterized by clusters and exhibiting only a few nodes that have a large number of links. While the scale-free network is appealing in the clustering characteristics of its nodes, it does not appear to accurately capture nodes with $k < 3$ and exhibits a singularity at $k = 0$ (isolates).

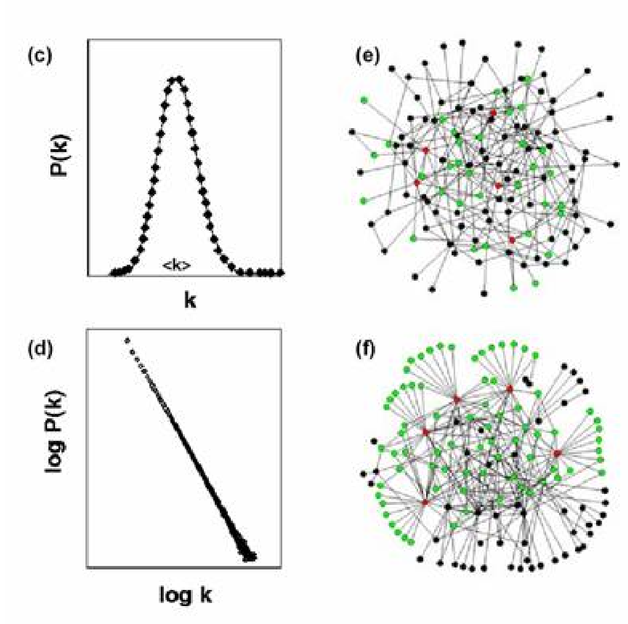


Figure 1: Copied from Barabási et al. [7]. Randomly generated network (c) and (e); scale-free network (d) and (f).

A major challenge is that few data- based examples of social networks from large groups (>50) exist to review and compare.[1]. Coleman [4] reported complete reciprocal friendships for eight friendship networks from four high schools, boys and girls separately. The networks ranged from 146 to 230 students. Coleman did not report isolates. Furthermore, his data only appears to be scale-free only above $k = 4$. In a more recent paper, Moody [8] described a high school social network with dense local clusters and sparse connections between subgroups in the network. Moody, however, did not appear to check the network for scale-free characteristic.

Klovdahl [13] researched social networks in an urban area (Canberra, Australia) by interviewing 183 randomly selected persons who nominated over 6,000 associates. The random walk strategy in data collection provided a unique approach in gathering social network data, albeit with unproven results. While their network appeared to exhibit some small world attributes, they did not observe that the networks were scale-free. Somewhat to the contrary, the distance between two individuals (shortest graph-theoretic path) ranged from 1 to 17. Furthermore, only 33% of the individuals were separated by six steps or fewer ('six degrees of separation') as would be suggested by scale-free networks.

In short, neither scale-free nor random networks provide an obvious advantage over the other.

3 Gang/Terrorist Analogy

As we probed the terrorist literature for information that would allow both the construction of the social architectures and population of the required data sets, we found that data on 'real' terrorist groups

[1] There are several reasons for this lack. This type of data is very time-consuming to collect. Furthermore, if collected, there was no way to efficiently manipulate it prior to the development of sophisticated social network codes. Finally, amount of data that needs to be manipulated for very large groups once all actual and potential connections are developed is beyond the capacity of most existing statistics packages.

was lacking. This is for obvious reasons given the secret nature of terrorist organization and the bias in data that has been collected in a potentially coercive manner. Members of terrorist groups are not easily accessible to researchers as are members of other types of social groups. Indeed, the conduct of the research itself could take on political overtones (e.g. to gain the confidence of the subjects, the researcher often has to [appear to] endorse the group's goals and methods). As a result, the small amount of available data is itself suspect (such as interviews conducted with prisoners), and much of the writing on terrorist groups and on terrorism itself is secondary, based on others' data.

To overcome this shortfall, we decided to use U.S. urban street gangs as an analog for terrorist groups. An initial focus on street gangs would help us create a basic architecture that could be determined to be reasonably reliable in that the principles on which the architecture was constructed would be determined reflect the target that is being simulated and the model could be validated to some degree against available data. We could then translate that basic architecture into a terrorist environment, elaborating and changing as necessary.

We chose U.S. urban street gangs for several reasons. First, there is a great deal of data available on them [15][16]. They have been studied for many years and from a variety of perspectives. Perhaps most importantly for the recruitment and growth questions we wanted to ask about terrorist groups, there are several studies that compare non-gang versus gang members of at-risk youth identifing qualities that result in gang membership and those deter at-risk youth from membership [15]. These studies also explored causal correlations among various social factors and gang membership. Second, we believe that individuals join gangs for many of the same reasons that are posited for recruits to terrorist organizations. The youths (for they generally are youths) are marginalized both socially and emotionally in some way [2], and are seeking some mechanism for self-realization and self-actualization that the dominant culture does not afford them. The gang provides a strong social identity and sense of acceptance [20]. Through actions in the gang, the individual can exert influence on the world in ways he could not otherwise thereby gaining access to various types of resources. However, in order to exert that influence, someone must pay attention. Hence the public nature of both terrorist and gang activities, the concept of 'violence as theatre.' In short, we believe that while the analogy between street gangs and terrorist groups is not perfect, it is a strong enough for us to proceed.

4 Recruitment Model Development

4.1 Vision

The motivation behind the underlying architecture for the Seldon toolkit takes into consideration a varied collection of issues including the incorporation of social science theory and data in the model, current capabilities of agent-based simulation toolkits, and how to potentially advance agent-based simulation in the future. These issues provide input into the Seldon architecture design, which is part of the vision and development for the recruitment model based on how groups/organizations form.

The hybrid design of the Seldon architecture attempts to extend the traditional ant-like behavior of current agent-based simulation models. Figure 2 illustrates the three different levels currently proposed for the agent-based architecture. Level 1 of the architecture is where the simple agents

[2] It is important to note that such marginalization is not necessarily through poverty. It can occur through cultural isolation or marginalization, such as is found with many of the ethnic gangs—and, may argue, the expatriated (actual and potential) terrorists.

(a.k.a. adolescent males), which share some similarity to the ant-like agent design, reside. The simple agents reside in a neighborhood that includes a school locale.

Level 2 of the architecture introduces a unique capability known as abstract agents. Unlike the simple agent the abstract agent provides the user a software entity for representing social or institutional concepts in an abstract manner. While the extended vision will give the abstract agents the ability to interact with each other and simple agents, the current architecture only permits interactions with the simple agent entities. In the current Seldon architecture there are two abstract agents: the School and the Gang.

Level 3 of the architecture remains only a vision for future advancements in the Seldon architecture beyond this current LDRD. The cognitive model(s) that reside at the third level of the architecture are an attempt to push current agent-based simulations that do not hold state information. The idea behind the cognitive model addition was to permit the agents at Level's 1 and 2 to one day tap into this higher-level model. The cognitive model would provide the simple agents with history to provide better understanding of social behavior.

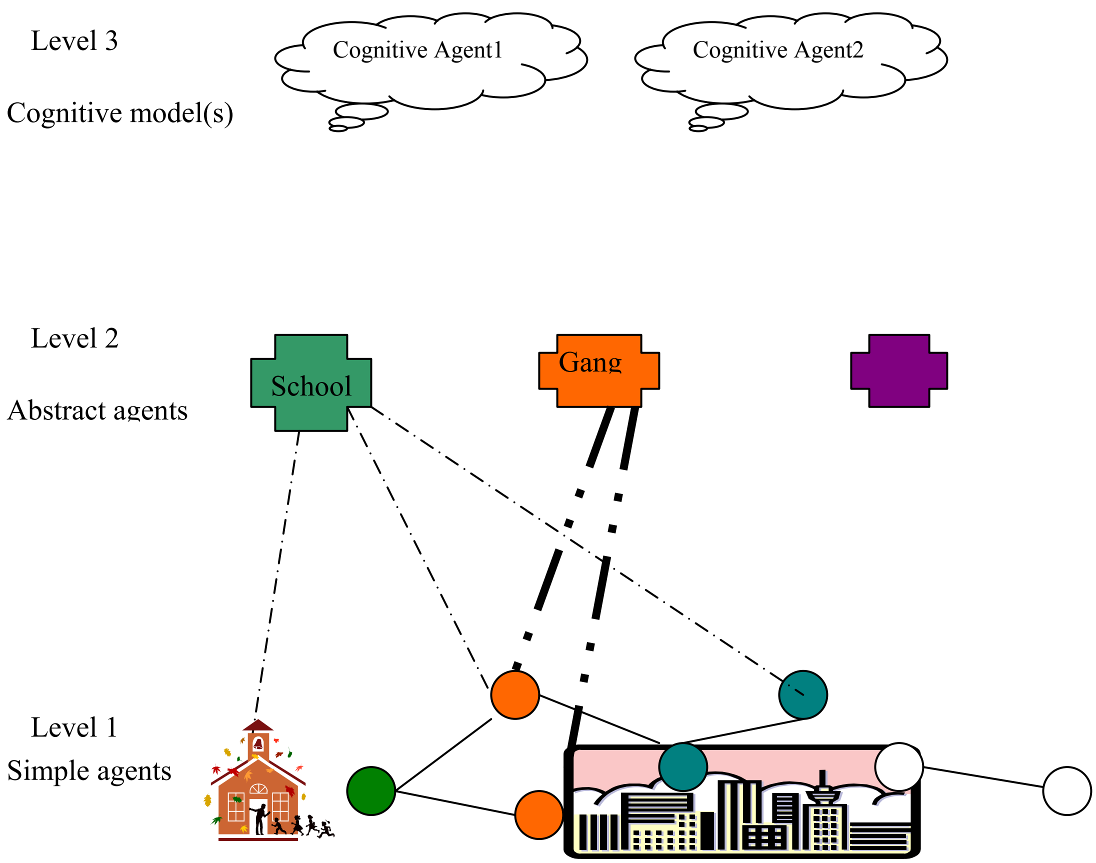


Figure 2: Seldon architecture

Providing cross-references for the different levels of the architecture are the interactions between and within the different levels. Within Level 1, these interactions are represented as straight lines and dotted lines between Level 1 and Level 2. These interactions are used to capture the different networking that drives the emerging social behavior of the underlying model. Note that the simple agents are not fully connected and that the abstract agents are not fully connected to the simple agents at Level 1. The mixing of interactions provides greater flexibility through, reflecting a varied collection of behaviors in any social system.

4.2 Development

The model architecture is represented in Figure 2. The simulation takes place in physical community of a neighborhood and a school building, which do not overlap.

4.2.1 Agents Defined

Two types of agents have been defined for this model, adolescent teen males and an abstract agent, “School”. Teen males were chosen because the availability of urban gang data and research on this demographic subsegment [15][16]. To be explicit, no other types of individuals, like females, exist.

Each of the teen agents is described by three attributes:

- Attribute 1
- Attribute 2
- School Attendance Tendency, or SAT (Hi/Lo)

The first two attributes 1 and 2 have not been explicitly defined herein, but think of descriptors like religion, socioeconomic background, ethnicity, or hair color. The purpose of the attributes is for the generation of network links that are dependent on homophily, or the general similarity between any two agents. While we have somewhat arbitrarily allowed for only two attributes, the number of descriptors for each agent can be virtually limitless. With binary descriptors for the two attributes, we effectively have four agent categories: 00, 01, 10, and 11. If one were to extend the list of attributes, one could imagine a DNA-like binary string that can uniquely describe each agent.

The last descriptor, SAT, is simply the tendency of any agent to go to school on any given day. For simplicity, we have divided the population into two tiers: Hi and Lo SAT. A distribution of school attendance tendencies could also have been implemented and might be considered in future. Hi SAT attends school regularly (frequencies are user-defined), whereas the Lo SAT more likely skip school.

The abstract agent, “School”, is a new concept used to capture an institution representing a collection of individual agents, that is more than an aggregate of the individuals. In other words, the whole is greater than a sum of its parts. With such a uniquely defined abstract agent, we can then capture the non-linear dynamics between individual and group. In other words, the individuals are each influencing the group while the group is influencing each individual.

The abstract agent, “School”, represents an aggregate of all the positive influences from attending school such as education, teacher role models, family investment, and self-esteem building activities. Studies have correlated the positive influence of school attendance on the academic functioning which result in a reduction of gang activity. [14]

4.2.2 Time

The concept of time, and time increments, within a social simulation is difficult to map into real time. In this project, we are interested in group dynamics occurring typically over the time span of months. To that end, we initially chose a time increment of one day. The frequency of processes, like agent-agent interactions per time step, was then selected to be consistent with a one-day increment.

The current Seldon model defines the following:

- Adolescent agents spend the entire day (365 days per year) either at school or in the neighborhood, mutually exclusive.
- No after school
- No weekends
- No summers
- Adolescent agents do not age

4.2.3 A Day in the "Life" of a Teen Agent

The day-to-day "life" cycle of the Teen Agent is represented in the Figure 3. The agents interact with other agents every day, the frequency of which is user-defined. For each time step, an agent must first decide whether or not to attend school that day. The decision to attend school is stochastic and biased by a user-defined probability. For example, a user might assign a probability of 70% to Hi SAT agents for school attendance. For each day, there is 70% likelihood that it will attend school. The agents do not maintain an attendance history and so are not affected by their past attendance decisions.

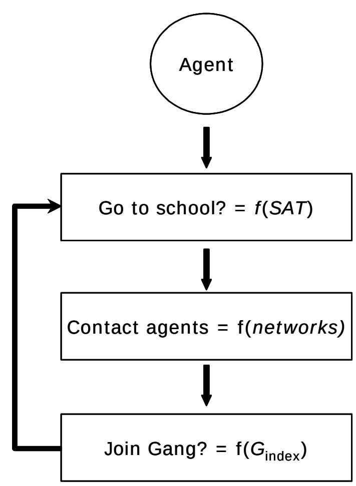


Figure 3:Day in the Life of a Teen Agent

After deciding whether it will attend school, the agent can only interact with the agent cohort that has made the same decision. Truants can therefore only interact with truants and the same holds true for school-attendees. Furthermore, all agents are restricted to interacting with only agents that are directly linked to them in their social network. Recall that all agents in the gang network are linked, as are all agents in the school network, non-gang network, and out of school network.

The model randomly selects which agents will interact with a user-defined bias toward different social networks. For example, interactions can be biased to favor interactions between friends over schoolmates.

During an agent-to-agent interaction, one agent can recruit or dissuade the other agent from joining a gang, and vice versa. The strength of the recruitment or dissuasion is user-defined and tracked per individual agent through the Gang Index, or G_{index}, defined as

$$G_{index} = w_{gang} N_{gang} - w_{non-gang} N_{non-gang} - w_{school} N_{school} \tag{1}$$

where:

w	=	Weights
N_{gang}	=	Number of contacts with a gang agent
$N_{non\text{-}gang}$	=	Number of contacts with a non-gang agent
N_{school}	=	Number of days of school attendance

Equation (1) mathematically represents the opposing forces in the dynamics of gang recruitment. On one side, the gang members are influencing the non-gang agents to join the gang. On the other side, non-gang agents and the “school” are dissuading agents from joining gangs and possibly leaving them if they have already joined.

G_{index} is tracked for each agent and represents the propensity of an agent to join a gang. The index is cumulative. At the end of every day, each agent decides whether or not to join a gang. The decision hinges exclusively on whether or not G_{index} exceeding a user-defined threshold. Once exceeded, an agent joins the gang. However, the agent also leaves the gang when their G_{index} drops below the threshold. The model does provide for a sticking mechanism for staying in a gang, i.e. for the influence of the agent’s past association with the gang.

5 Software Toolkit Development

5.1 Implementation Motivation

The Seldon toolkit contains modeling and simulation software that is designed to study the formation of human organizational structures. The main objective of the software implementation was to develop an agent-based social simulation that models extreme social dynamic transitions such as the tipping point phenomena of gang or terrorist emergence. To expedite this process the software team decided to survey several agent-based simulation packages instead of creating software from scratch. The team used the set of criteria listed in Table 1 to evaluate the different agent-based software packages. The first six items in the Table 1 are the most important criteria, while the last two items are not as significant during the exploratory development of the Seldon toolkit. The ability to scale is significant since several types of agent toolkits exist; however, not all are capable of supporting agent-based simulation. Unlike traditional multi-agent systems, that tend to support hundreds of software agents, agent-based simulations need to support thousands of agents.

Table 1: Agent Toolkit Criteria

Criteria	Pros and Cons
Language – Java	Portability, web applet.
Scalability – number and type of agents.	From 10^2 to 10^5 of agents. Should support finer grain agent development and not simply heavy weight agents.
Plug 'n Play – upgrade.	Plug in new modules easily (i.e., visualization tools, cognitive model).
Debugging – developer and support.	Get support for creators on bug fixes. Debug the model as you develop it, what is going on in the model.
Visualization – some graphic support.	What types of visualization software exist if any?
Communication/message passing – some support.	Future need for some agents that can communicate directly to other agents.
Emergent behavior – support capability.	Pack should support light weight reasoning ability and not just heavy/course grain agents.

After evaluating several software packages the team selected the University of Chicago's Social Science Research Group's RePast software package, which is a framework for creating agent-based simulations [9]. RePast, which is an acronym for REcursive Porous Agent Simulation Toolkit, was an almost ideal development environment for Seldon's purposes. It provides plenty of debugging and visualization support, is Java based, and is well used by the social simulation community. The RePast package is a library containing Java 1.4 classes for creating, running, collecting data and visualizing the results of the agent simulation. RePast borrows much from the Swarm [10] simulation toolkit and can properly be termed "Swarm-like." In addition, RePast includes such features as run-time model manipulation via GUI widgets.

While RePast proved to be a good starting point for the exploratory phase, the team felt RePast had limitations and began to modify the underlying RePast library. These modifications later resulted in the creation of a Seldon Java library, which provides classes for creating a model (i.e., simple agents, abstract agents), running the simulation with a hybrid scheduling mechanism (based on lessons learned from RePast), graphical user interface (GUI), and visualization software.

The Seldon software toolkit depicted in Figure 4 is written in Java JDK 1.4.2 and consists of two main components (in Java these are known as packages), the *"model" – simulation model* and *"graphical user interface"* (GUI) – user interface to the model initialization and visualization. These two components are integrated by a third component that acts as the interface between the model and the GUI, known as the *Main program*. This separation in the flow of data provides the developer and user with multiple execution and parallel development methods for the toolkit. Therefore, changes can be made to the model and test without having to run the complete graphical interface, which proved to be time consuming. The toolkit also has two supporting components the "*visualization*" – display simulation data and the *"data"* – initialization data and results from the simulation.

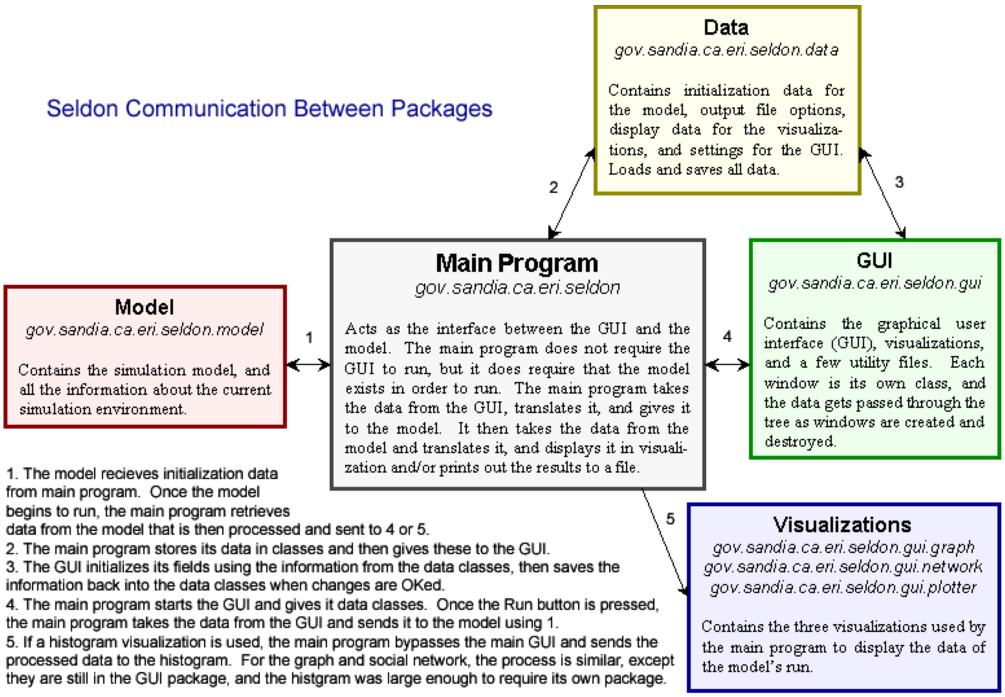


Figure 4: Seldon integrated software structure

5.2 Components

5.2.1 Graphical User Interface (GUI)

The GUI is the main mechanism used to capture user provided input data and visualize the results of the current simulation run. While this section provides detailed overview of the GUI, the visualization tools will be covered in detail in section 5.2.2. The main window is displayed in (a) and provides the user with the four menu options: (1) File, (2) Simulation, (3) Visualizations, and (4) Help. Each menu extends into a collection of subtasks that can be performed by the user of the Seldon toolkit.

Figure 5(1), the **File** option let's the user *Open* an existing configuration file to run the simulation or *Save* the active simulation configurations to a file. The *Output Options* subtask lets the user save the simulation data to a text file using different delimiter options. This text file could later be used as input to another software package like Microsoft Excel™ to do additional analysis.

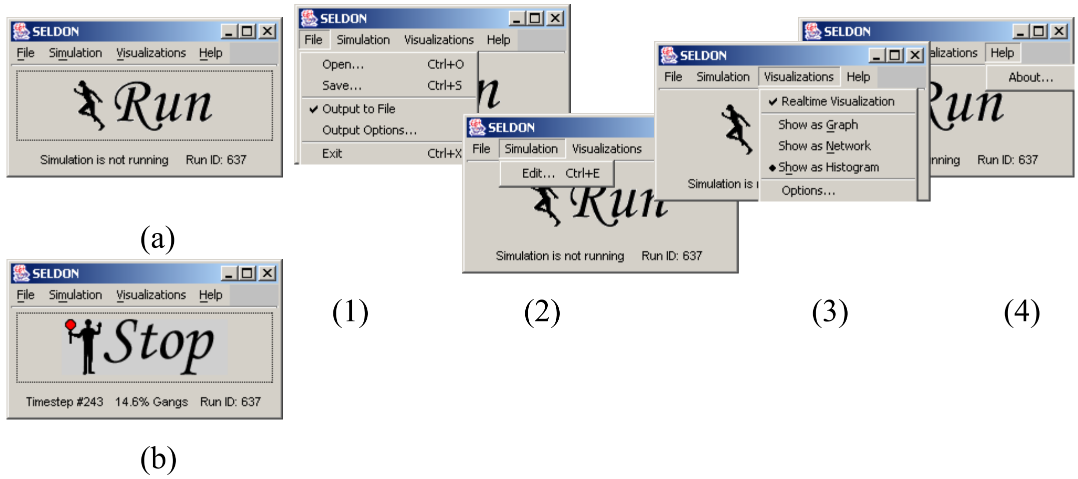


Figure 5: Seldon start up window and menu options

Figure 5(2), the **Simulation** option allows the user to modify the settings used to initialize and run the agent simulation. A detailed discussion of the model inputs is covered in section 5.6 on model configuration. Figure 5(3), the **Visualizations** menu will permit the user to select from a variety of visualization software packages. See section 5.2.2 for detailed information about the visualization package. Figure 5(4), the **Help** menu currently only contains the Seldon logo and information about the leveraged software packages. Additional help is provided to the user in the form of mouse-over pointers that bring up additional windows informing the user of useful information throughout the simulation.

Once the user has either loaded or created input for the simulation he or she is ready to run the simulation by pushing the **Run** section of the window. While the Seldon simulation is running the main screen will resemble Figure 5(b). This screen also provides the user with feedback on the number of timesteps and the percentage of the overall agent population that has become gang members. The Run ID simply provides an ID for each run that is saved as part of the output file.

5.2.2 Visualization Tools

The complexity and large number of entities associated with agent-based simulations traditionally has required the incorporation of different visualization techniques. These visualization packages extend the traditional GUI providing graphing and analysis of results in a real-time or post-processing manner. The research and complexity behind the creation of this visualization software leads many social simulation developers to incorporate pre-existing software.

The Seldon software toolkit leverages three existing visualization software codes illustrated in Figure 6. First the user determines if the visualization will occur in real-time or post-processing. Next the user can select one of the three software graphic packages to display. For each package there are various options that can also be set to adjust the format of the selected simulation data. The three visualization packages provided in Seldon include line graphs, social network graphs, and histograms (e.g. multiple views histogram or 3-D histogram)

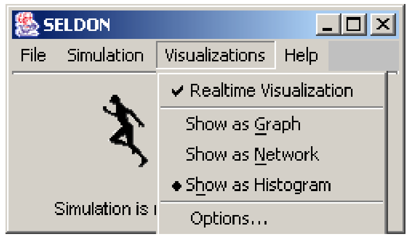


Figure 6: Seldon visualization selection screen

The line graph software is based on the 2-D data plotter and histogram tool, known as PtPlot from UC Berkeley [11]. The PtPlot software is a part of the [Ptolemy II]() project, which comes from the homogeneous modeling and design group. PtPlot is used to display a live updating graph of the data being processed by the Seldon model. Figure 7, provides an illustration of the output of this line graph software. The user has the option of setting different parameters and ranges for the line plots. These include selecting the data set to be plotted: (1) entire population or (2) gang or non-gang, or (3) high or low attendees, or (4) number of relationships (fixed number or range). The user may also select a subset of settings that permits him or her to vary the same values provided in 2-4 for the entire population. The x-axis displays the number of time steps in the simulation and the y-axis displays the percentage of gang growth.

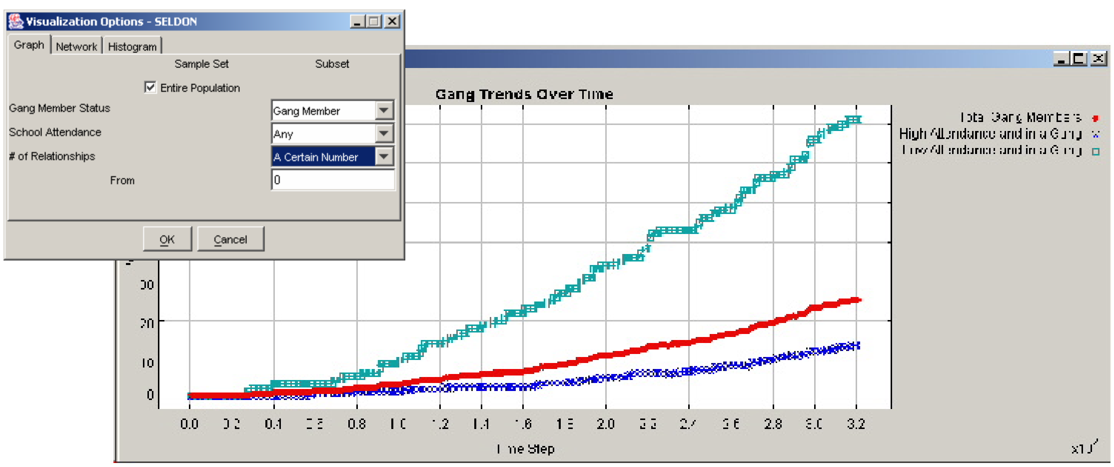


Figure 7: PtPlot line graph

The social network software provides the user with a graph depiction of the connectivity between the different agents based on their friendship interactions. This package will let the user display a combination of up to three categories each based on color, shape, and size respectively. Figure 8 illustrates the social network software option window in the upper left corner and a resulting graph from this simulation run. This figure yields high (squares) and low (circles) attendance of each agent by the shapes of the node/leaf on the graph. The color scale provides knowledge concerning the number of gang and non-gang members as they enter or leave their perspective groups.

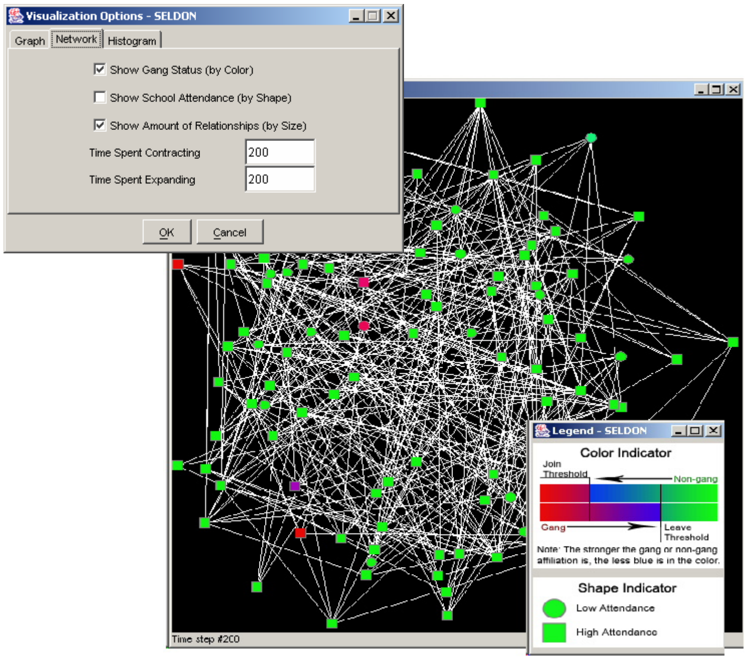
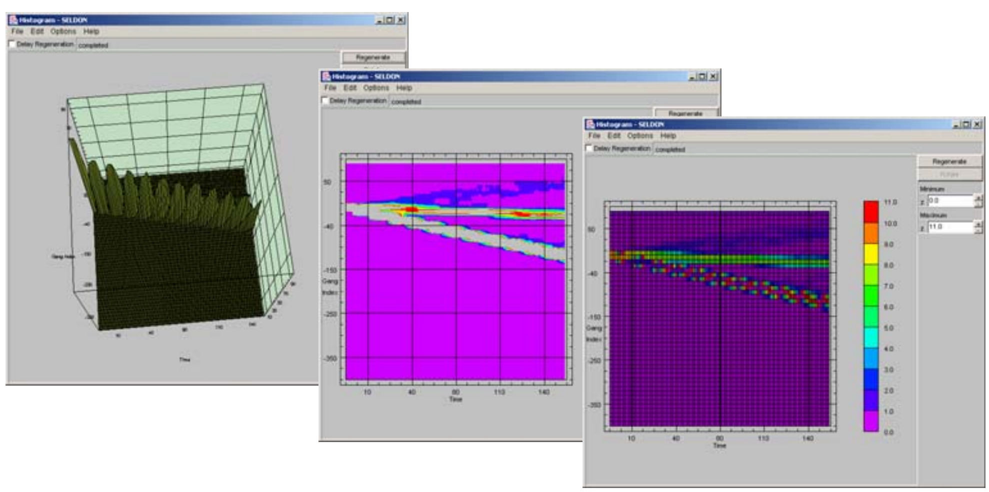


Figure 8: Social network

The plotter software is a Java(tm) applet which plots surfaces defined by explicit two-variable mathematical function (i.e: z = f(x,y)). Yanto Suryono developed the original plotter software known as Surface Plotter, in 1996 [12]. The surface plotter software was adapted for use in the Seldon toolkit with alterations to fit the data set. The plotter software can display the three different types of histogram plots shown in Figure 9: (a) surface, (b) contour, and (c) density.



(a) (b) (c)

Figure 9: Surface plot examples

While the visualization packages used in the phase-1 implementation of the Seldon toolkit provides the user with basic analysis capabilities, these packages will not meet the enhanced needs of the future development. Missing from current visualization is the ability to display multiple types of graphs concurrently. For instance, it would be valuable to see a surface plot while watching the line graph(s). The social network is not a real-time mechanism that responds to changes in the social network(s), it is a fixed entity that simply illustrates a snap-shot of the friendship network. Changes in this network over time cannot be captured with the current software.

5.2.3 Agents

The major components of the Seldon toolkit are the agents that represent several different aspects of the underlying model. The philosophies behind the agent-based models are their ability to represent behavior(s) of individual agent. Once the agents populate the model it can be run, thereby exhibiting emergence of societal level behavior based on the interactions of the individual agents. The agent-based approach used in the Seldon toolkit is a unique hybrid design (see section 4.1for more details) that uses two categories of agents, whose interactions provide relationships that directly reflect social organizational issues. The first category of agents, called the simple agents, are basic entities representing people. The second category of agent resides in Level 2 of the Seldon hybrid design, and are known as the abstract agents. The abstract category of agents is used to introduce conceptual characteristics into the agent-based simulation model.

The largest numbers of agents, the simple agents (SA), represent the adolescent males in this recruitment model. While this model provides the user the ability to represent individuals, these representations are based on categories of adolescent males. There is no attempt in this model to provide a degree of granularity to actually represent an actual adolescent male, i.e. Joe or Tom. In equation (2) SA is a fixed collection of characteristics used to represent the adolescent inner city male. These characteristics include parameters for: high school attendance and low school attendance. The sa_p represents the parameter lists from both the set of adolescent and/or interactions that can be manipulated by the inputs from the users. The interaction parameters are defined by the social networking structure of the model and are described in Section 5.2.4. A complete list of all model parameters is provided in Appendix 8.1.

$$\begin{aligned} &SA \equiv simple\ agent \equiv (adolescent\ male\ characteristics) \\ &\forall\ SA\ \exists sa_p\ \ where\ \ p \ni \{\{adolescent\ \ parameter\}, \{\mathrm{int}\, eractions\}\} \end{aligned} \tag{2}$$

The abstract agent (AA) gives the model a way to capture conceptual aspects of the recruitment model. By turning these concepts into agents we are providing a mechanism to promote a different type of interaction between these abstractions and the simple agents. The AA are manipulated with a set of parameters like the simple agent entities. Each aa_p comes from a set of gang and/or school parameters that represent social interactions like thresholds for joining and leaving a gang. In the recruitment model there are two abstract entities the *Gang* and the *School* [14].

$$AA \equiv abstract\ agent \equiv (social\ concept;\ institutional\ concept)$$
$$\forall\ AA\ \exists aa_p\ \ where\ \ p \ni \{\{gang\ \ parameter\}, \{school\ \ parameter\}\} \qquad (3)$$

The *Gang* abstraction represents the concept(s) behind a gang of inner city adolescence, or what aspects coordinate with the adoption of the gangster lifestyle. It is important to note that we model only gang recruitment, and our model has neither gang-violence, turf struggle, competition among gangs, nor drug sales. That is, we are concerned only with an 'attractiveness' factor of the gang, not with its activities *per se*. The *School* abstraction provides the concept(s) associated with a traditional school place where adolescents congregate. The *School* abstraction also represents the opposing influence (i.e., teachers, after-school activities, etc.) to the Gang abstraction.

The abstract agents also determine the structure of the social networking (see section 5.2.4 for details) for the Seldon recruitment model. This approach permits us to introduce some potential organizational control over the social interactions of the agents within the model.

5.2.4 Social Networks

The incorporation of networking into agent-based simulation was initially based on graph theory and systems dynamics models derived from physics. While early non-social network approaches provided interesting results they did not provide a realistic correspondence to existing social human systems. Today some researchers are beginning to incorporate social networks into their agent-based simulation models. The social networks provide a mechanism to study the influence of underlying social structures on organizational behavior, such as recruitment and gang formation.

The current Seldon toolkit generates a single social network, known as the *friendship network* [19]. Simple agents in the *friendship network* can interact; however since the network is not a fully connected graph the agents cannot interact with everyone. To promote additional interactions the *friendship network* is further refined with internal clustering that represents different types of relationships or influences between the simple agents beyond the concept of friendship. Within the friendship network the interactions between the agents are governed by their similarities, which is based on the combined view of Attributes 1 and 2 (see Section 4.2.1 for discussion), shown in Table 2.

Table 2: Attribute Combinations

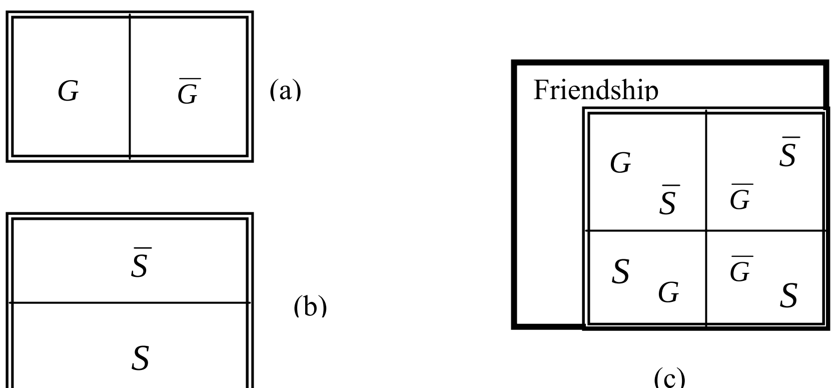
Attributes	Attribute Combinations			
Attribute 1	0	1	0	1
Attribute 2	0	0	1	1

The two internal clusters identified for the gang recruitment model are for "schoolmate" and "gang" associations. The "schoolmate" cluster consists of a collection of simple agents that are currently all attending school for the day. The interactions between the simple agents in the "schoolmate" cluster are influenced by the school surroundings. The "gang" cluster is composed of simple agents that are members of the "gang" during that time-step. The concepts associated with the gang influences the interactions between the agents in the "gang" cluster.

The implementation behind the *friendship network* and the refining internal clusters are based on the three set representations listed in equation (4). For each regular set there is also a complementary negative set. Since all simple agents are members of some or all of the three main sets, it holds that the *friendship network* is also composed of a set of sets. This fact is illustrated in the last equation in (4).

$$\begin{aligned}
&G \equiv \{in\ gang\} && \overline{G} \equiv \{not\ \ in\ \ gang\} \\
&S \equiv \{in\ school\} && \overline{S} \equiv \{not\ in\ school\} \\
&F \equiv \{friends\} && \overline{F} \equiv \{not\ \ friends\} \\
&Friendship\ network \equiv \{F, \overline{F}, G, \overline{G}, S, \overline{S}\}
\end{aligned} \tag{4}$$

The relationship between the collection of simple agents (SA) and the network and clusters are illustrated in equation (4). This equation bring to light the fact that individual SA may belong to both internal clusters at the same time. By refining the friendship network with the two internal clustering techniques we are able to enhance and provide some unique social control over the simple agent population. Figure 10 illustrates how we implemented these three interaction categories as an integrated collective. In Figure 10(a) and (b) illustrate the further subdivision of the internal clusters shown in equation (5). When we superimpose these two clusters the resulting combination is shown in Figure 10(c), which we combine with the larger *friendship network.*



$$\forall sa \in \left(\left(G \vee \overline{G}\right) \wedge \left(S \vee \overline{S}\right)\right) \wedge \left(F\right) \tag{5}$$

Figure 10: Social network refinement for schoolmate and gang

The current implementation of the social networking strategy in the Seldon recruitment model provides some unique insight into how socially structured networks affect the formation of organizations like gangs. While the internal clustering provided a means to refine the larger friendship network, we would replace the current implementation with real social networks in any future advancement of the Seldon toolkit.

5.2.5 Scheduling Mechanism

The Seldon toolkit has a scheduling mechanism that drives the overall simulation of the recruitment model. The motivation behind the stepwise scheduling mechanism was described earlier in Sections 4.2.2 and 4.2.3. Each time step within the model is equal to 1 day, which is implemented as two scheduling phases for a single time step in the simulation. Many changes within the model happen during a given time step and this section provides an overview of some of the decision making processes the simple agents does during a given day.

Phase-1of the time step randomly selects one simple agent (a.k.a. adolescent male) from the list of unselected simple agents. Each selected simple agent must decide if they are going to school. This decision alters a few different parameters, including: school attendance and gang-index. Whether a simple agent goes to school is a function of their school attendance, and a random variable. If a simple agent goes to school, the school reduces their gang-index. Going through all of the agents in phase-1 should take O(n log n) in the worst case.

The decision made in Phase-1 directly determines what types of social interactions are possible in future implementations of the Seldon toolkit. Once at school, the simple agent may not contact simple agents outside of the school (and vice versa). However the simple agent may contact other simple agents from friendship network, "schoolmate" cluster, and "gang" cluster. If the simple agent is not in school they may contact other simple agents in the friendship network and "gang" cluster. The decision to contact/interact with another simple agent is a function of the matching between the network or clusters shared between the two agents.

Phase 2 of the scheduling begins with randomly selecting simple agents to determine what interactions they will have with other simple agents. The decision made in Phase-1 directly determines what types of social interactions are possible in Phase-2. The abstract agents also provide additional input during these interactions based on the social parameters setting during each Phase-2 step. Once a simple agent has decided whom to contact, these simple agents exert influence on each other. Members of "gang" cluster will increase the gang-index of other simple agents they contact (or are contacted by), and non-gang members will either decrease or not influence the gang-index of other those with whom they come into contact.

At the end of Phase-2 each simple agent must 'decide' if he will alter his gang affiliation based on a gang threshold provided by the user during the initialization process. The decision is straightforward. A simple agent joins the gang if its gang index exceeds the threshold and leaves the gang otherwise. At the end of each time step, the orders of the simple agents are shuffled to promote the randomization of selections for the next time step. A simple agent's decision to go to school is independent of the state of the world, and therefore the order of the simple agents for this phase is irrelevant.

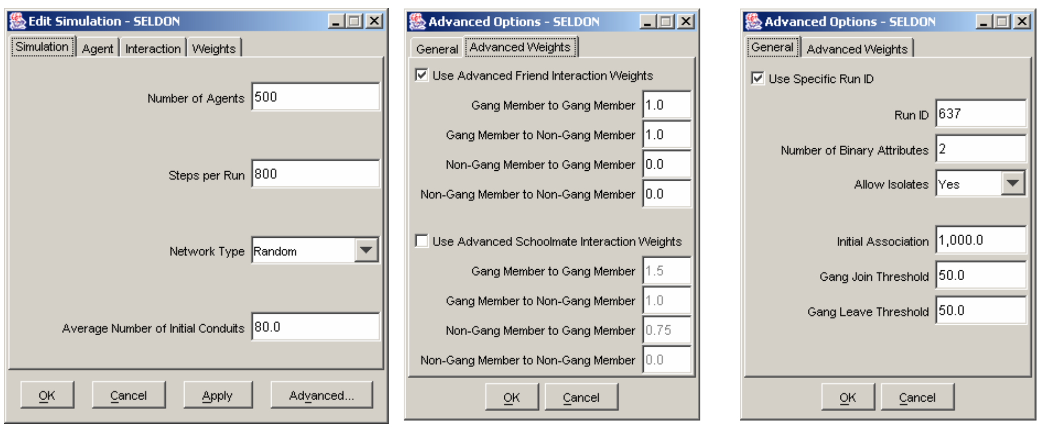
5.3 Model Configuration

5.3.1 Interactive model configuration

The Seldon GUI provides the user with the ability to *only* manipulate the parameters used to configure the underlying gang recruitment model. This means that the GUI panel code will have to be changed to reflect model development that is not directly related to the methodology of recruitment. The current gang recruitment model can be modified by selecting the **Edit** menu under the Seldon run/stop window as depicted in Figure 5(2) in Section 5.2.1.

Figure 11(a), provides a snapshot of the edit simulation secondary window that is displayed when the Edit option is selected. The edit simulation window provides the user with four main categories of parameters settings Simulation, Agent, Interactions, and Weights each under their own tab. By selecting the **Advanced** button on the right lower corner of the edit simulation window, the user will be given two additional parameter category options as seen in Figure 11(b): Advanced Weights and (c) General.

The significant parameters under the simulation tab [Figure 11(a)] are the number of agents (e.g., male adolescent) in the run and the number of steps or days (a.k.a. time steps) the simulation should run before it stops. The current recruitment model only provides a random networking option (see Section 5.2.4 for details on the current social networks implemented in this toolkit). The advanced weights tab [Figure 11(b)] contains parameters that affect the interactions between the different types of agents (gang or non-gang) based on their internal network connections. If these are not selected the system will default to the non-gang/non-gang values.

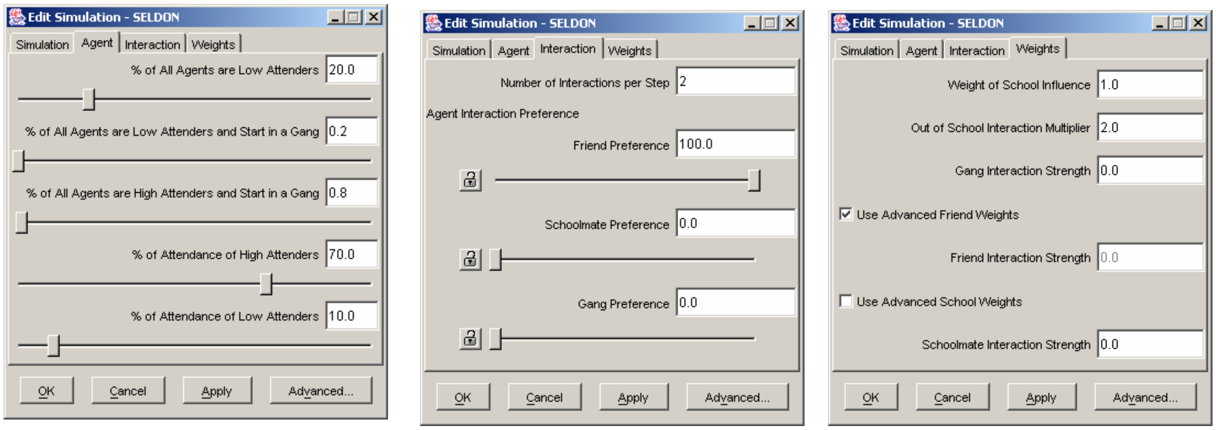


(a) **(b)** **(c)**

Figure 11: Seldon parameter setting windows

The general [Figure 11(c)] tab will permit the user to set general parameters that apply to each agent. The significant parameters include the number of attributes for each agent; the current gang recruitment model has been tested consistently with 2 binary attributes. The user can also determine the threshold levels each agent will use to determine when he joins/leaves a gang, respectively.

The screen shot for the remaining edit simulation tabs are located in Figure 12(a) – Agent, 6(b) – Interaction, and 6(c) – Weight. Each of these areas permits the user to manipulate different parameters dealing with overall population settings, interactions preferences, and strengths of influences respectively. To promote variability in the parameter mix Figure 12(a) and (b) uses slider mechanism to select their values. The settings applied in Figure 12(a) directly reflect the population breakdown for the initial gang recruitment run. By altering these parameters the user is able to manipulate the school attendance factors and view the effects of school on the overall agent population.



(a) **(b)** **(c)**

Figure 12: Edit simulation parameter screen shots

The interactions window depicted in Figure 12(b) will let the user adjust the number of interactions per agent for each step. The user can also adjust the interaction preferences or types along with the level of interactions the agents have between each other. The slides are co-dependent, which translates into mixed control between the friend and schoolmate preferences. The higher the friend interactions the less significant the schoolmate preferences are rated and vice versa. Similar relationships exist between the gang preference and the friend/schoolmate preferences. As the gang preference increases the other two are forced to be less significant in the simulation.

Figure 12(c) provides a collection of basic/general weights that can be set by the user to manipulate the influences of school and interactions between school, gangs, schoolmates, and friends. The user can also bypass the general friendship and school weight in favor of the more extensive advanced weights shown in Figure 11(b).

5.3.2 Configuring the model with a file

The Seldon toolkit also permits the user to provide, a configuration file as input to the simulation. An example of this configuration file can be found in Appendix 8.2. The configuration file can be generated from scratch or generated from saved settings provided from a prior simulation session. All of the parameters used in the configuration file correspond directly to the parameters from the interactive GUI. Note that (randomSeed = runID) in the configuration file.

5.3.3 Initializing the model

Once the input parameters to the model are provided the initialization process begins by setting up the simple agents, which are fixed for the remainder of the simulation. Figure 13 provides a pictorial flow of the simple agents and their breakdown into the four different categories. The user specifies the percentage of different agents in these categories as input parameters. The four categories of simple agents reflect and enforce the granularity of the model by not permitting the user to identify or manipulate individual people at any lower level.

Each simple agent is initialized with a starting gang-index. This is set at 0 for a non-gang member and user defined for non-gang members. The effects of other user-supplied parameters are seen in the percentage of high or low school attendance, which is fixed throughout the simulation. However the gang membership of the simple agent can vary during the simulation.

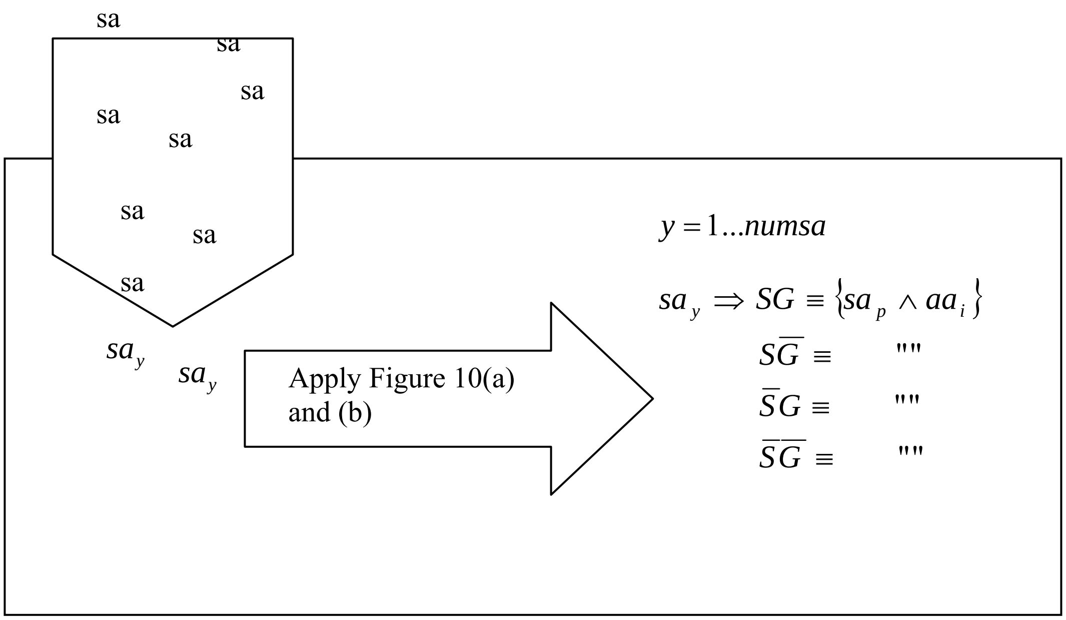


Figure 13: Initialization process for simple agents in the model

6 Simulation Results and Discussions

The values for the model parameters were estimated initially through conversations with social science domain experts. They were then tuned by adjusting the parameters so they provided outputs that appeared reasonable to experts. Although we have allowed for bias in agent interactions, the work this year focused exclusively on the friendship network. In short, agents can only interact with other agents wit which they have a direct friendship link.

Figure 14 shows the emergence of gangs in a simulation that is initially populated with 1% gang agents (5 of 500 agents). The average number of friendship links per agents is 5 links per agent, which is a social network density of ~1% (5/499). The Lo Attendees are more quickly converted to gang members, as they do not have the daily positive influences from the abstract school agent. However, even the Hi Attendees are also eventually converted to gang members in this scenario. With a relatively low network density, interactions between any two agents are repeated frequently since the potential interactions with different agents are small (average of 5). Recruitment therefore follows systematically and sequentially.

The distribution of friendship links based on our network generation algorithm is shown in Figure 15. The actual characteristics of a social network are the topic of current debate. Several researchers argue that some social networks follow power-law distributions [2] rather than the normal distribution generated by random graph theory. However, friendship networks have simply not been studied in great detail. It is therefore dubious that either random-graph or power-law distributions provides an accurate representation of friendship networks [3].

Early data from Coleman [4] shows a reciprocal friendship data for eight sets – four high schools, boys and girls separate – and plotted in Figure 15. At higher degrees of connection (links), a power-law distribution appears applicable. However, the power-law break-down occurs for dyads (one link) and isolates. Coleman does not even report isolates, or loners.

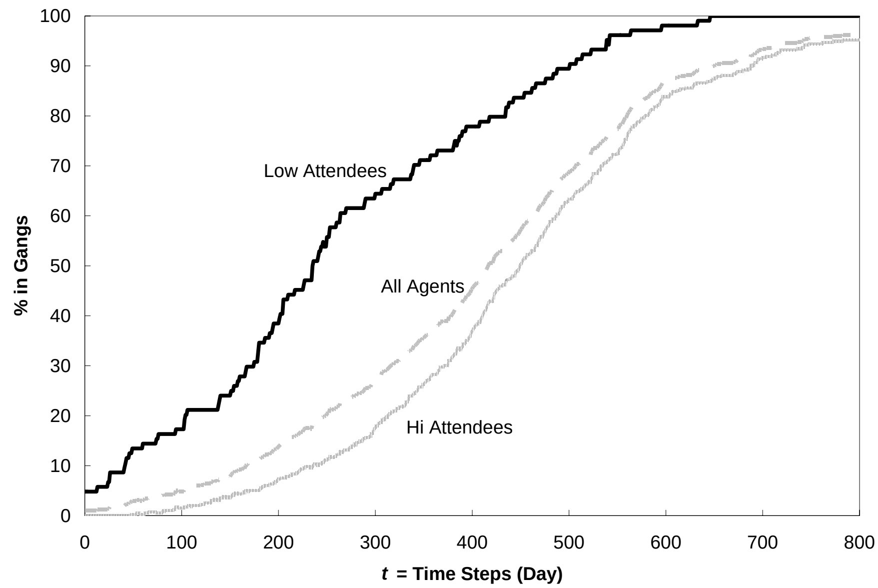


Figure 14: Growth of Gangs with initial network density of 1%

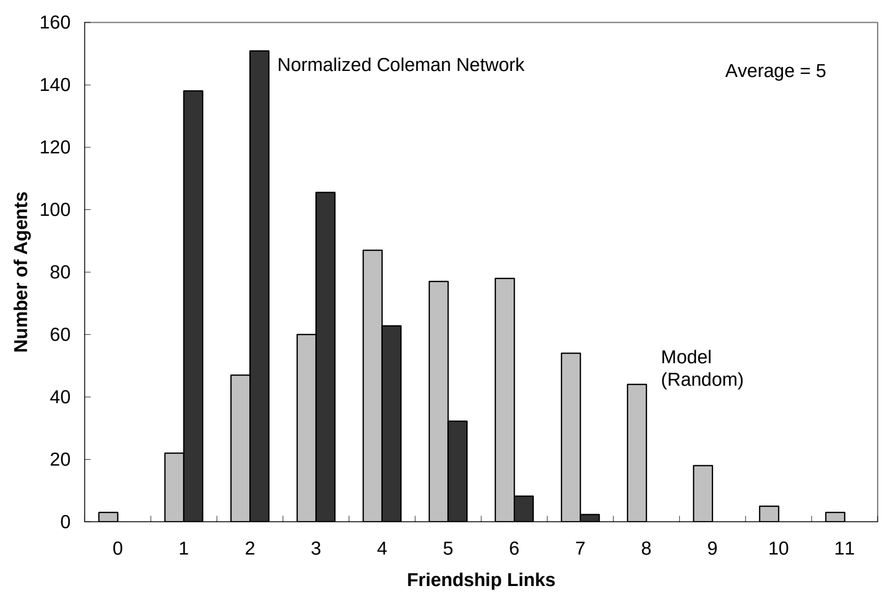


Figure 15: Histograph of the distribution of friendship links (average = 5)

6.1.1 Effect of Network Density on Gang Growth

The effect of network density on the gang growth rate is shown in Figure 16. Increasing network density is akin to increasing the number of different agent-agent interactions and the distribution of links is shown in Figure 17. Each teen agent has the potential of interacting with a greater number of different agents. The influence of one agent on another is now diffused as the likelihood of repeated interactions is decreased, as the choice of agent-agent interactions is completely random. The possibility of a recruiter agent that seeks to interact with the same agent(s) repeatedly has not been implemented in this model, although recommended for future versions.

The emergence of gangs at $\rho_{\text{network}} = 0.16$ proceeds in a distinctly different pattern than it does at $\rho_{\text{network}} = 0.01$, a greater number of Figure 14 to Figure 16. At the higher ρ_{network}, the gang membership has two distinct growth spurts, or tipping points, rather than the gradual increase at the lower ρ_{network}. Interestingly, the system in Figure 16 also exhibits two metastable regions as well, t = 0 to 80 and 200 to 300. During the metastable periods, the system does not appear to be changing, as measured by the gang membership, an observable characteristic. Yet, the system is moving toward a tipping point.

The first period of metastability (t = 0 to 80) appears to mitigate gang growth by numbers in that repeated interactions between any two agents are statistically unlikely. In contrast, the second period of metastability (t = 200 to 300) is mitigated by physical segregation. During the second period of stability, Lo Attendees have all been converted to gang members and Hi Attendees are all non-gang members. There is some daily mixing between the two populations although it is physically limited; however, there is sufficient mixing to ultimately tip the Hi Attendees to joining gangs.

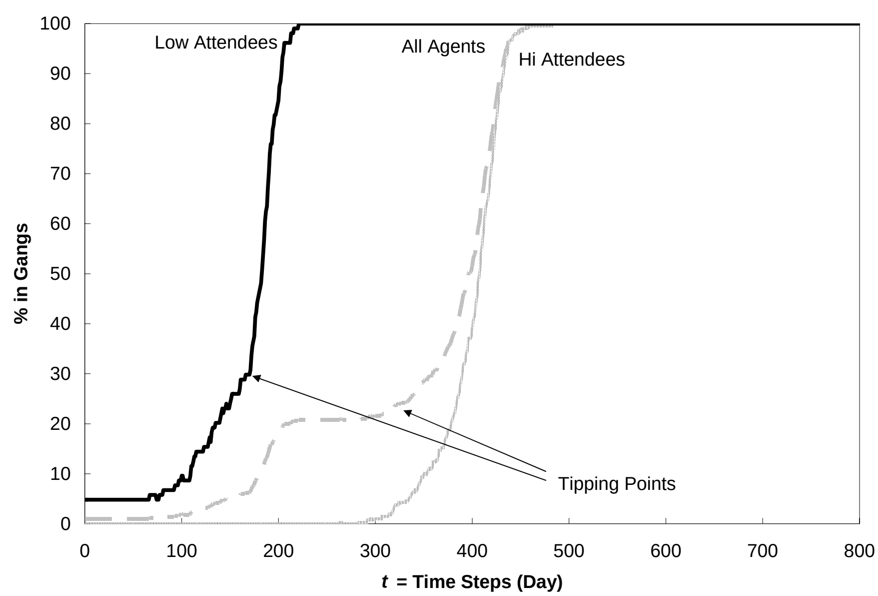


Figure 16: Growth of Gangs with initial network density of 16%

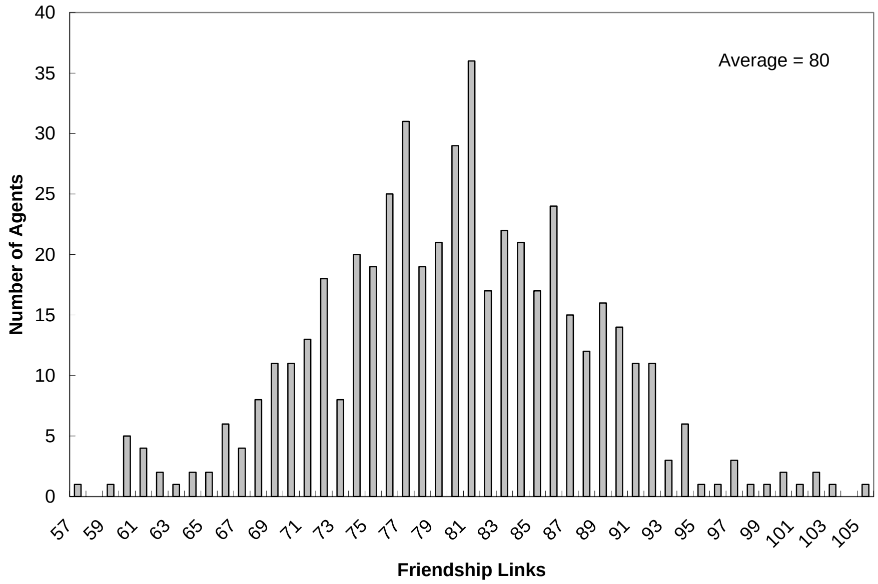


Figure 17: Distribution of links with a network density of 16%.

We can get a much better understanding of the system dynamics by dis-aggregating the data and look at the individual agents. Figure 18 represents the same simulation as depicted in Figure 14, at $\rho_{network} = 0.01$, but plotting G_{index} instead of Gang%. G_{index} provides an indicator of an agent's proclivity to join a gang, whereas Gang% only indicates if an agent has joined. The histograph shows the number of agents at each G_{index} as a function of time. Two separate trajectories are evident in Figure 18. The subgroup of Hi Attendees moves steadily away from gangs (lower G_{index}) as its members attend school daily. Lo Attendees also move away from gangs, but less strongly as their school attendance is not as high (10% vs 70%). In both cases, the infiltration of gang agents through the network is slow and steady with first the conversion of Lo Attendees and eventually Hi Attendees. There appears to be a resilient 3% that do not convert.

The histograph at $\rho_{network} = 0.16$ is even more revealing as shown in Figure 19. The behavior of the Lo and Hi Attendees are depicted by two distinct trajectories. The Lo Attendees convert first as before. However, we can now see the movement of the G_{index}. During metastable periods when gang membership is not growing, the G_{index} is generally increasing indicating a societal move toward gangs. The cause of the tipping points becomes clearer in Figure 19. At t ~ 180, the Lo Attendees move to gangs, as measured by G_{index}, starts exceeding the threshold (50), gang conversion is autocatalyzed. As more agents are converted to gang members, the system tips as gang members beget gang members and the group of Lo Attendees convert virtually instantaneously into gang members. The effect cascades into the Hi Attendee subgroup almost immediately as evidenced by the change in trajectory, although the observable effect (gang membership) is not seen until t ~ 300.

During this second metastable period (t = 200 to 300), the conversion of Hi Attendees to gang members is mitigated both by the physical separation from Lo Attendees, who are mostly gang members, and their history of school attendance leading to low G_{index}. However, the mitigation is fleeting as Hi Attendees begins tipping to gang membership as soon as the first Hi Attendee becomes a gang member, at t ~ 300.

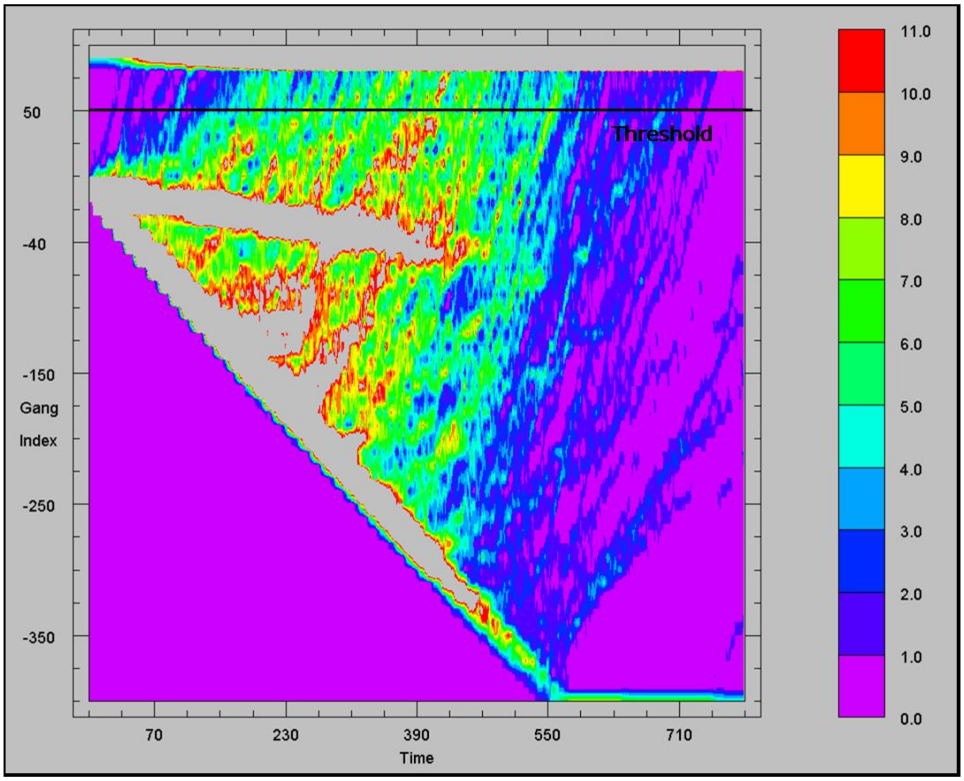


Figure 18: Histograph of gang growth with initial network density of 1%

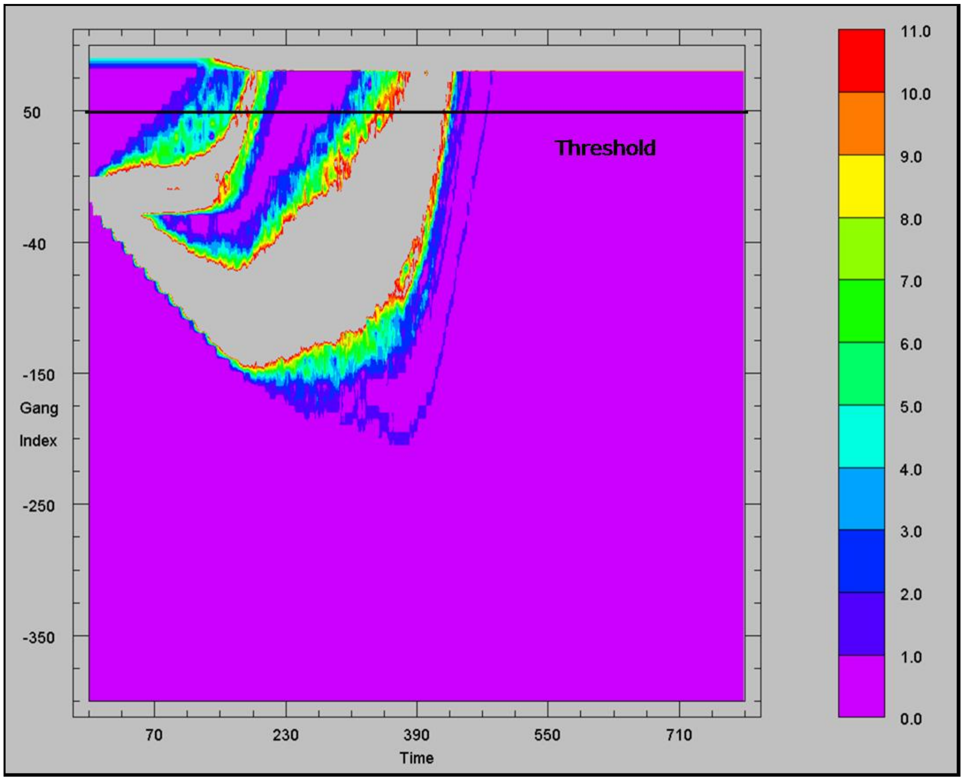


Figure 19: Histograph of gang growth with initial network density of 1%

6.1.2 Effect of increasing school attendance on gang growth

One frequently proposed top-down policy to mitigating the growth and activity of gangs is to increase the school attendance of those agents who are the otherwise chronic truants, the Lo Attendees. As with all model scenario analyses, no programs or techniques are suggested or implied for improving the school attendance. However, the model does provide insights into the system response for gang membership from increasing school attendance.

Increasing school attendance from 10 to 30% for Lo Attendees ($\rho_{network} = 0.01$) decreased gang membership negligibly from 97% to 89%. A further increase in attendance from 30 to 50% subsequently decreased the gang membership more significantly from 89% to 46%, with still a substantial portion of the agent population converting to gang members.

In contrast, increasing the school attendance from 10 to 30% for a network density of $\rho_{network} = 0.16$ dramatically decreased the gang membership from 100% to 4%. Since we do not know the effective social network densities in a high school, we can only speculate herein that a dual approach of increasing both attendance and interactions among agents may be a more effective approach than either approach alone.

6.1.3 Effect of non-gang agents disrecruiting other agents on gang growth

In simulations reported up to this section, agent-agent influence has been unidirectional in that gang agents influence other gang and non-gang agents. Non-gangs exerted no influence on other agents. The only positive influence (away from gangs) was provided by school attendance. We now ask the following question: How does positive influence from other non-gang agents impact the simulation output? We answer this by increasing the weight (influence) of non-gang agents on both non-gang and gang agents (parameters are "friendInteractionNonGangGang" and "friendInteractionNonGangNonGang"). Previously, these parameters had been set to null (0) and the weight (influence) of gang agents had been set to unity (1). We increased the non-gang weights from 0 to 0.2, effectively allowing the non-gang agent influence to be 20% that of gang agent influence.

This interdiction was very effective for system with $\rho_{\text{network}} = 0.01$ as the final gang membership decreased from 96% to 7%. Even more effective was the interdiction with a higher network density of $\rho_{\text{network}} = 0.16$ where there were no gangs in the final population.

6.1.4 Effect of "improving school effectiveness" on gang growth

The abstract agent, school, embodies all positive attributes of school attendance in dissuading agents to joining gangs, including but not limited to elements such as a path to higher education and better paying jobs and improved self-esteem from participation in school activities. We do not attempt to identify or define the factors that contribute to 'school effectiveness' in this model. "Improving school effectiveness" also cannot therefore be precisely defined herein. From the purely computational perspective, we define "improving school effectiveness" as increasing the weight (influence) of school's positive impact on agents every day that they chose to attend school.

Our simulations indicated that increasing the school effectiveness by 50% would subsequently decrease gang membership from 95-100% down to 30-40%, depending on the network density. The decrease is not inconsequential, but not as significant as some of the other interventions.

6.1.5 Effect of "anti-gang education" on gang growth

Agents in all simulations up to this section have been given identical threshold G_{index} for converting to gang members. We recognize that the natural distribution of such a threshold is highly variable. However, identical agent thresholds were a simplification for the initial model. Future iterations will allow for a distribution of thresholds more closely resembling that which we find in society. Intervention with anti-gang education is implemented here by increasing the G_{index} threshold for each agent.

Initially, the threshold was set to $G_{\text{index}} = 50$. We increased threshold to $G_{\text{index}} = 100$ and found that the gang growth was delayed by 2X, with either $\rho_{\text{network}} = 0.01$ or 0.16. In the lower density network ($\rho_{\text{network}} = 0.01$), the gang growth seemed to be slowed by 50%. In the higher density network ($\rho_{\text{network}} = 0.16$), both tipping points were still observed but their onsets were delayed by 2X. In short, the "anti-gang education" appeared to slow the growth of the gang, but did not appear to change the final observation that all agents became gang members. This introduces an interesting dimension to efforts to reduce urban street gangs. Can we simply slow the gang recruitment dynamics sufficiently so that the agents can leave the environment (graduate from high school) faster than their 'gang index', or proclivity to join a gang, reaches a critical threshold?

7 Conclusion and future work

7.1 Future Work

Immediate extensions to the Seldon toolkit will provide a variety of upgrades to the current implementation of the software. We specifically would like to upgrade or replace current implementations components for the GUI, visualization, and social networking component. While we will keep the current format of the Phase-1 Java applet GUI; however, we will update the underlying implementation by addition some automation the GUI creation. The current Phase-1 GUI was hard coded with the parameters from the recruitment model. This code will be removed and replaced by a routine that will automatically update the GUI by reading a file containing the parameters for the given model. To accomplish this task the current implementation of the parameters for the recruitment model were moved to promote more flexibility.

The current visualization software is based on available open source software from different universities. The team opted to use these packages because visualization for complex agent-based simulations is an open research topic with no fixed solutions. This software has proven to be problematic and not supported by the third party developers. We will replace the three different open sources packages with commercial software that will permit us the ability to provide real-time simulation of multiple graphic windows.

The social networking component for the Phase-1 software did not correctly implement the social networking, thereby limiting developers to a single social network model. We will replace the current social networking model to permit multiple networks and a limited set of interactions between these networks. This addition will make the Seldon model unique in its inclusion of multiple social networks. Most agent-based social simulations do not include any social networking or only one social network. This modification will place the toolkit into un-chartered territory, to reduce the risk of this modification we will control the creation and types of interactions between the social networks.

Beyond these immediate modifications future advancements to the Seldon toolkit should include the incorporation of a cognitive model and/or agent to complete the hybrid agent-based architecture proposed in Section 4.1. The incorporation of the cognitive model is perhaps the riskiest component of the agent-based vision, due to the lack of supportive research in this area. We believe a cognitive agent will provide advanced capabilities for the simple agents in the current model. These capabilities would include a collective world model; history and knowledge of past events; and physical localization of simple agents. Additionally an advanced cognitive model could provide emotions and learning/adaptability for a collection of agents with minimal behaviors. Another future advancement is the investigation of scale-free and random networks for the recruitment model. In the initial investigation we did not note any difference in the results of these networks; however, further investigation is required.

7.2 Conclusion

The Phase-1 development of the Seldon toolkit successfully provides users with a tool to hypothesis and test group recruitment for inner city gangs. While the Phase-1 toolkit dealt explicitly with U.S. inner city gangs the same recruitment model can be used to study other organizational recruitment issues. In FY04 a brief LDRD is extending the Phase-1 recruitment model to align the model with Middle Eastern terrorist recruitment in a European setting. These enhancements illustrate the latitude of the underlying hybrid agent architecture that is flexible enough to support the recruitment concept across lateral domains.

The hybrid architecture used in this research provides a unique integration of technology and concepts from the interdisciplinary fields of agent-based modeling, social science, simulation, and cognitive sciences. This architecture differs from traditional computational social dynamic simulations because of its (1) multi-level design; (2) abstract agent representing social or institutional concepts; and (3) cognitive model/agent. While this hybrid architecture has only been implemented through the Level-2 state it has already illustrated interesting results for recruitment model studies. In additional to these unique factor this architecture also provides social interactions based on social network models. This capability is unique to only a handful of existing agent-based simulation toolkits.

There are many lessons that have been learned during the development of the Seldon toolkit. We believe important foundations have been established between teams of social scientist and software developers to generate a preliminary tool for analyzing gang recruitment. While many additional years (10-15) of development and research will be needed to fully realize the hybrid agent-based architecture presented in this document, we believe this body of research is off to a good start.

8 Appendix

8.1 Model Parameters

The model parameters are defined and summarized in the following table.

Table 3 Descriptions of Model Parameters

Parameter	Description
numAgents	Number of agents
numBinaryAttributes	Number of attributes per agent
numInitialContacts	Ave # links to other agents in friendship network
contactsPerStep	Number of interactions per time step for each agent
useAdvancedFriendWeights	Weight for friendship network
friendInteractionFrequency	Frequency between 0 and 100. One of three networks, all of which must add up to 100.
friendInteractionGangGang	Gang-to-gang weight, within friends network
friendInteractionGangNonGang	Gang-to-nongang weight, within friend network
friendInteractionNonGangGang	Nongang-to-gang weight, within friend network
friendInteractionNonGangNonGang	Nongang-to-nongang weight, within friend network
loGangProb	% of initial gang agents who are low attendees
hiGangProb	% of initial gang agents who are hi attendees
gangAssociation	Initial gang index of starting gang agents
gangJoinThresh	Threshold of gang index for joining gang
gangLeaveThresh	Threshold of gang index for joining gang
gangInteractionFrequency	See "friendInteractionFrequency"
gangInteractionWeight	Weight for friendship network
lowAttenderProb	Fraction of initial agents who are low attendees
lowAttenderAttendance	Probability a low attendee goes to school each day
highAttenderAttendance	Probability a high attendee goes to school each day
schoolWeight	Weight for school network
useAdvancedSchoolmateWeights	0 for No, 1 for Yes (GUI only)
schoolmateInteractionFrequency	See "friendInteractionFrequency"
schoolmateInteractionGangGang	Gang-to-gang weight, within school network
schoolmateInteractionGangNonGang	Gang-to-nongang weight, within school network
schoolmateInteractionNonGangGang	Nongang-to-gang weight, within school network
schoolmateInteractionNonGangNonGang	Nongang-to-nongang weight, within school network
stepsPerRun	Time steps per run
outOfSchoolMultiplier	Weight multiplier for interactions out of school

8.2 Configuration file

```
randomSeed 637
randomLock 1
numAgents 500
numBinaryAttributes 2
numInitialContacts 6.0
contactsPerStep 2.0
useAdvancedFriendWeights 1
friendInteractionFrequency 100.0
friendInteractionGangGang 1.0
friendInteractionGangNonGang 1.0
friendInteractionNonGangGang 0.0
friendInteractionNonGangNonGang 0.0
loGangProb 0.0020
hiGangProb 0.0080
gangAssociation 1000.0
gangJoinThresh 50.0
gangLeaveThresh 50.0
gangInteractionFrequency 0.0
gangInteractionWeight 0.0
lowAttenderProb 0.2
lowAttenderAttendance 0.1
highAttenderAttendance 0.7
schoolWeight 1.0
useAdvancedSchoolmateWeights 0
schoolmateInteractionFrequency 0.0
schoolmateInteractionGangGang 1.5
schoolmateInteractionGangNonGang 1.0
schoolmateInteractionNonGangGang 0.75
schoolmateInteractionNonGangNonGang 0.0
stepsPerRun 800
```

9 References

[1] M. Macy and R. Willey, From Forms to Actors: Computational Sociology and Agent-based Modeling, Annual Review of Sociology, 2002. Vol. 28:143-166.

[2] Réka Albert and Albert-László Barabási, Statistical mechanics of complex networks, Reviews of Modern Physics, January 2002, 74, pp 47-97.

[3] Kathleen Carley, personal communication, 2003.

[4] James S. Coleman, The Adolescent Society – The Social Life of the Teenager and its Impact on Eduction. The Free Press of Glencoe, New York, 1962.

[5] Duncan J. Watts, Six Degrees, W.W.Norton & Company, Inc., NY, 2003

[6] M.E.J. Newman, The Structure and Function of Complex Networks,

[7] Albert-László Barabási, Zoltán Dezső, Erzsé Ravasz, Soon-Hyung Yook, and Zoltán Oltvia, Scale-free and hierarchical structures in complex networks, November 25, 2002.

[8] James Moody, Peer influence groups: identifying dense clusters in large networks. Social Networks, 23, 2001, 261-283.

[9] Repast, http://sourceforge.net/projects/repast/

[10] Swarm, http://www.swarm.org/

[11] PtPlot, http://ptolemy.eecs.berkeley.edu/java/ptplot5.2/ptolemy/plot/doc/index.htm

[12] Surfaceplot, http://www.fedu.uec.ac.jp/~yanto/java/surface/

[13] A.S. Klovdahl, Social Networks in Contemporary Societies, unpublished report, 2002.

[14] F.H.G. Gilyard, The Competition between Gangs and Schools, PhD Dissertation, The Fielding Institute, 1999.

[15] Dukes RL, Martinez RO, Stein JA. Precursors and consequences of membership in youth gangs. Youth & Society, 1997; 29(2): 139-165. [S&D]

[16] Huff R. Criminal behavior of gangs in Aurora and Denver, Colorado, and Broward County, [producer], 1995. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2000.

[17] Levine S. Youth in terroristic groups, gangs, and cults: The allure, the animus, and the alienation. Psychiatric Annals, 1999; 29: 6.

[18] Moen P, Elder, G & Luscher K Examining Lives in Context: Perspectives on the Ecology of Human Development, 1995.

[19] Stokman and Zeggelink 1996. 'Self-organizing' friendship networks. In W.B.G. Liebrand and D.M. Messick (eds), *Frontiers in social dilemmas research.* Berlin: Springer

[20] Moore J. Isolation and stigmatization in the development of an underclass: The case of Chicano gangs in East Los Angeles. Social Problems, 1985, 33:1.

Distribution

5	MS9217	Ben Wu	16000
1	MS0839	Gerry Yonas	16000
1	MS0839	Tim Moy	16000
1	MS0839	Jessica Turnley	16000
1	MS0839	Curtis Johnson	16000
1	MS0839	Judy Moore	16000
1	MS0839	John Whitley	16000
1	MS0839	Rick Craft	16000
1	MS0839	Nancy Hayden	16000
1	MS0839	Tom Karas	16000
1	MS0839	Peter Merkle	16000
1	MS0839	Ron Pate	16000
1	MS0839	Ken Miller	16000
1	MS0839	Patricia Gingrich	16000
1	MS1188	Vivian George	15311
1	MS0830	Ann Speed	12335
1	MS1188	Mike Bernard	15311
1	MS1109	Rich Pryor	9216
1	MS0318	Dianne Barton	9209
1	MS0127	Bob Floran	12111
1	MS9003	Ken Washington	8900
1	MS9003	James Handrock	8960
1	MS9003	Chuck Hartwig	8940
1	MS9011	Barry Hess	8941
1	MS9019	Brian Maxwell	8945
1	MS9019	Steve Carpenter	8945
1	MS9037	Jim Berry	8947
1	MS9012	Steve Gray	8949
1	MS9915	Mike Koszykowski	8961
1	MS9217	Steve Thomas	8962
1	MS9012	Jerry Friesen	8963
1	MS9007	Mike Hardwick	8964
1	MS1170	Russell Skocypec	15300
1	MS1188	John Wagner	15311
1	MS1188	Chris Forsythe	15311
1	MS1188	Steve Tucker	15311
3	MS9915	Nina Berry	8961
1	MS9915	Teresa Ko	8961

1	MS0323	D. Chavez, LDRD Office, 1011	
3	MS9018	Central Technical File	8945-1
1	MS0899	Technical Library	9616
1	MS9021	Classification Office, 8511, for Technical Library, MS 0899, 9616	
		DOE/OSTI via URL	