

Simulación computacional de grupos de investigación en la Universidad Estatal a Distancia de Costa Rica

Andrés Segura Castillo

Universidad Estatal a Distancia, 474-2050 San Pedro Montes de Oca, San José, Costa Rica; asegura@cenat.ac.cr

Colaboratorio Nacional de Computación Avanzada, Centro Nacional de Alta Tecnología, 1174-1200 Pavas, San José, Costa Rica

Recibido 19-IX-2009 Corregido 8-X-2009 Aceptado 13-X-2009

ABSTRACT

Computational simulation of research groups at Costa Rican Distance Education University. Social network analysis deals with quantitative studies of human groups and has established a set of measurements to understand their structure and behaviour. Nevertheless, their dynamism is only expressed to a limited extent and multiagent modeling and simulation has become a complementary analytical method. It enables researchers to simulate human interactions on virtual environments and to discover emergent patterns of behavior. This article presents a multiagent system that models and simulates the dynamics of research groups at the Costa Rican Distance Education University. The results of a first scenario point out that the initial amount of available resources in the system does not have a causal relationship with the clustering coefficient, i.e. the metric used to measure the researchers' tendency to form groups. The concept of resource must be reconsidered to include infrastructure and equipment availability. A second scenario shows that as the number of researchers increases, the clustering coefficient decreases. This effect is attributed to a tendency to equilibrium in the relative costs and benefits of maintaining a given number of researchers. More variables are needed for a reliable simulation.

KEY WORDS

Social simulation, research groups, multiagent systems, social network analysis.

RESUMEN

La modelación y simulación multiagente en el contexto de las ciencias sociales se ha convertido en un método de análisis complementario para investigadores de estas disciplinas, ya que facultan la creación de modelos computacionales que simulan las interacciones humanas en ambientes virtuales, con el fin de explorar la aparición de patrones de comportamiento emergente no observables a simple vista en el fenómeno real. En este estudio se presenta un sistema multiagente creado para modelar y simular la dinámica de los grupos de investigación de la Universidad Estatal a Distancia. Los resultados de un primer escenario indican que la proporción inicial de recursos disponibles en el sistema no tiene una relación causal directa con el coeficiente de agrupamiento, métrica que mide la tendencia que tienen los investigadores a formar grupos. Se plantea que debe reestructurarse el concepto de recursos en el modelo, con el fin de incluir no solo variables estrictamente económicas, sino también de infraestructura y disponibilidad de equipo. Por otra parte, en un segundo escenario fue posible establecer que el aumento en la cifra de investigadores es un factor que disminuye el coeficiente de agrupamiento, ya que el sistema trata de balancear los costos y beneficios de mantener una cantidad dada de los mismos. La investigación insiste que el modelo computacional creado no es un producto acabado, por el contrario, debe afinarse para incluir otras variables y aproximarse en mayor medida al fenómeno real.

PALABRAS CLAVE

Simulación social, grupos de investigación, sistemas multiagente, análisis de redes sociales

En general, las organizaciones existen como una estrategia colectiva para superar las limitaciones de las acciones individuales (Carley & Gasser 1999). Por ejemplo, en el contexto de la investigación científica, los investigadores se agrupan con el objetivo de incrementar sus posibilidades éxito. El estudio de este tipo de fenómenos de agrupamiento ha sido abordado tradicionalmente

mediante técnicas cuantitativas en ciencias sociales, de manera particular por el campo del Análisis de Redes Sociales (ARS). Según tal enfoque, las redes sociales están compuestas por relaciones estructurales, definidas como patrones regulares de interacción entre entidades concretas, por ejemplo, personas, grupos pequeños, organizaciones, etc. (Knocke & Yang 2008). Desde esta perspectiva,

es posible establecer una serie de mediciones o “métricas” que describen el comportamiento y la estructura de las interacciones en cuestión, sin embargo, en muchas ocasiones no son capaces de captar el dinamismo de los grupos humanos en general.

Debido a dicha limitación, el uso de métodos y modelos computacionales en ciencias sociales ha sido reconocido como un recurso que complementa los análisis cuantitativos tradicionales (Gilbert 2008). En particular, un Sistema Multiagente (SM), es decir, un sistema compuesto por múltiples agentes (componentes) de software y/o hardware autónomos capaces de interactuar en ambientes y tareas predefinidos (Lees et al. 2007), representa una herramienta de modelación social sin precedentes.

Según Oprea (2004), el supuesto que subyace esta afirmación, es el hecho de que, en la mayoría de los casos, este tipo de sistemas pueden entenderse como representaciones simplificadas de sociedades, donde el comportamiento de sus agentes o miembros, está definido por una serie de procesos, normalmente reglas, que se suponen existen en realidad, por ejemplo normas de interacción social (Oprea 2004). Además un SM brinda la ventaja de incluir variables que tradicionalmente no es posible contemplar en un modelo estrictamente estadístico, por ejemplo, la disponibilidad de recursos en un contexto dado, o los distintos roles que adoptan los agentes ante cambios en el ambiente (Fortino et al. 2006).

Así, es posible entonces diseñar un SM que refleje una situación social particular, es decir, que sea un modelo computacional (MC) del caso de estudio; con el objetivo de simular distintos escenarios y observar el comportamiento del sistema en dichas situaciones o “mundos” posibles (Troitzsch 2009).

Dicho enfoque simulacionista genera como resultado un conjunto considerable de datos, a partir de los cuales es posible inferir patrones de comportamiento emergentes en el MC. Las regularidades que surgen pueden ayudar tanto a una mejor descripción como comprensión de los fenómenos reales (Burger & Buskens 2009).

La presente investigación combina las métricas utilizadas en ARS y la capacidad de modelación y simulación multiagente para simular y analizar computacionalmente las interacciones y grupos que surgen entre investigadores de la Universidad Estatal a Distancia (UNED), específicamente, para establecer factores que propician o interfieren su funcionamiento. Esta propuesta resulta innovadora, ya que, el contexto de la investigación científica y su gestión, entendidas como ámbitos sociales propicios para la modelación y simulación multiagente, prácticamente no ha sido explorado hasta el momento.

METODOLOGÍA

La metodología de la presente investigación involucra el análisis estadístico propio del ARS y la modelación y simulación computacional mediante un SM, sin embargo, es importante aclarar cada uno de los pasos que este acercamiento requiere:

- Recolección y análisis de datos de entrada para el modelo.
- Diseño y construcción del MC.
- Verificación y validación del MC.
- Diseño y desarrollo de una serie de simulaciones para obtener un cuerpo de datos de salida estadísticamente significativo.
- Extracción de patrones de comportamiento emergente en el MC.

A continuación se describe los detalles de la metodología empleada.

Recolección y análisis de datos de entrada

Es importante mencionar que, en general, la cantidad y la calidad de los datos recolectados es la que permite encontrar las regularidades de comportamiento en las cuales insiste el ARS. El manejo adecuado de los mismos es indispensable para dar plausibilidad social al modelo, es decir, para que este cuente con un sustento que le permita simular un comportamiento semejante al fenómeno acaecido en la realidad.

Los insumos para un MC construido con un SM, pueden provenir de fuentes muy diversas, sin embargo, actualmente se reconoce como fuentes viables las siguientes (Hudlicka & Zacharias 2004):

- Literatura de archivos (historiales, memorias, documentos organizacionales, etc.).
- Análisis de tareas.
- Entrevistas.
- Estudios empíricos y estadísticos (informes, censos, encuestas, etc.)

En el caso de esta investigación se revisó la documentación de los archivos de la Vicerrectoría de Investigación (VI) de la UNED y se llevaron a cabo una serie de entrevistas con investigadores y gestores de dicha entidad para recolectar el cuerpo de datos necesario. Una vez finalizadas estas actividades se procedió a analizar los datos, con el objetivo de establecer mediante métodos estadísticos, las regularidades de comportamiento latentes en la muestra

recolectada. Los resultados de este proceso son expuestos en la sección correspondiente del presente artículo.

En este punto, es importante mencionar que en cualquier investigación empírica de redes sociales es necesario especificar los límites o alcance de la misma, es decir, cuáles son las variables que interesa analizar (Knoke & Yang 2008).

Para efectos del modelo de los grupos de investigación en el contexto de la UNED se consideran las siguientes variables:

- Cantidad de investigadores.
- Cantidad de recursos disponibles.
- Cantidad de integrantes en un grupo de investigación.
- Probabilidad de vínculo entre investigadores.
- Probabilidad de desaparición de un vínculo entre investigadores.

Además, para medir la influencia de dichas variables en los grupos de investigación de la UNED se calcula el coeficiente de agrupamiento (Watt & Strogatz 1998), el cual consiste en una métrica propia del ARS, que indica el grado en el que las entidades en un modelo tienden a agruparse.

Diseño y construcción del modelo computacional

Un MC mediante SM normalmente se desarrolla a partir de un modelo lógico del sistema, que establece las entidades relevantes en el sistema, sus acciones e interacciones (Louie & Carley 2008). Es indispensable, para efectos del MC, que la cantidad de variables sea, en la práctica, tratable computacionalmente, de manera que sea factible llevar a cabo las simulaciones en un tiempo y espacio razonables.

La especificación lógica de los grupos de investigación de la UNED fue extraída a partir del análisis de los datos recolectados en la primera fase y fueron afinados por el criterio experto de la vicerrectora. Así, las entidades modeladas corresponden a investigadores, cuyas características incluyen:

- *Un grado académico*: Consiste en un número entero que representa el título obtenido hasta el momento, por ejemplo 1 para bachillerato, 2 para licenciatura, etc.
- *Un costo operativo para la VI*: Se representa mediante un número entre 1 y 10 que indica el costo relativo del investigador para la VI, donde uno es muy poco costoso y 10 muy costoso.

- *Un beneficio operativo para la VI*: Esta característica denota el beneficio relativo que obtiene la VI de un investigador particular. Igualmente se establece una escala de 1 a 10, donde 1 es el menor grado de beneficio y 10 el mayor.

- *Un límite de investigadores con los cuales puede establecer vínculos*: Básicamente consiste en un número entre 3 y 7 que indica la cantidad máxima de miembros que un investigador tolera para formar grupos de investigación.

Seguidamente, se crean métodos computacionales, que aproximan, ya sea mediante heurísticas o fórmulas matemáticas, el comportamiento e interacciones de las entidades que integran el sistema a las situaciones reales. Este paso fue implementado en el lenguaje de programación Java y con el apoyo del conjunto de herramientas de modelación y simulación libres que provee REPAST (REPAST 2009).

Cabe resaltar que para efectos de la simulación multiagente todas las características de los investigadores modelados se inicializan aleatoriamente, según una distribución uniforme.

Verificación y validación del modelo computacional

El primer paso para verificar la correctitud del modelo, es llevar a cabo un análisis de sensibilidad (Gilbert & Troitzsch 2005), el cual consiste en una serie de simulaciones, cada una a su vez con una cantidad considerable de iteraciones, para corroborar que a pesar de los elementos aleatorios presentes, no existe un comportamiento caótico en ningún momento de la ejecución del MC. En la sección de resultados se incluye el análisis de sensibilidad correspondiente a la presente investigación.

En lo que respecta a la validación del MC se consideran dos niveles: el macro y el micro. El primero de ellos corrobora mediante criterio experto que las simulaciones cumplen con el comportamiento mínimo esperado, o sea, que se asemeja suficientemente al fenómeno real. A pesar de las críticas que despierta este primer nivel de validación, es importante aclarar que dada la naturaleza no estrictamente formalizable en términos matemáticos de los fenómenos sociales, el criterio experto es indispensable y en algunas ocasiones la única herramienta posible de validación (Louie & Carley 2008).

Ahora bien, para complementar la validación macro, se lleva a cabo un análisis a nivel micro, donde se emplean métodos estadísticos sobre un conjunto grande de datos resultantes de las simulaciones, proceso que, además de corroborar el comportamiento del MC, permite inferir

patrones de comportamiento emergente.

Diseño de las simulaciones y extracción de patrones emergentes

Para efectos de esta investigación se diseñaron dos escenarios de simulación con el fin de estimar la influencia que tienen las variables seleccionadas en el modelo, en particular, se hizo énfasis en la cantidad de recursos iniciales del sistema y en el número de investigadores que participan en el SM.

El primer escenario desarrollado plantea la variación de la cantidad de recursos disponibles en el sistema. Se realizaron un total de 200 simulaciones. Cada una contaba con 300 iteraciones y el valor de los recursos variaba, en orden ascendente, 100 unidades en cada simulación. Es importante mencionar que los recursos en el modelo representan rubros económicos y por el momento no contemplan otras posibles variantes como infraestructura u equipo.

Un segundo escenario consiste en la modificación sistemática de la cantidad de investigadores que participan de las simulaciones. En este caso se llevaron a cabo 200 experimentos con 300 iteraciones cada uno. La variación de la cantidad de investigadores fue de 50 agentes más en cada simulación con un valor inicial de 50.

Las simulaciones produjeron un conjunto de datos de salida que fueron analizados estadísticamente y sus resultados se presentan en el siguiente inciso.

RESULTADOS

Recolección de datos

A partir de los documentos de archivo de la VI y de 12 entrevistas llevadas a cabo con distintos investigadores de áreas diversas y gestores en la UNED se establecieron los siguientes patrones de comportamiento en los grupos de investigación:

- Hay mayor probabilidad de iniciar un grupo con colegas conocidos previamente.
- Hay mayor probabilidad de iniciar un grupo si existen suficientes recursos disponibles.
- Una vez establecidos, existe poca probabilidad de que los grupos se desintegren con el paso del tiempo (Fig. 1).
- Los grupos de investigación varían de 3 a 7 personas.

Es importante mencionar que las entrevistas realizadas contaban con una serie de preguntas que además de esta-

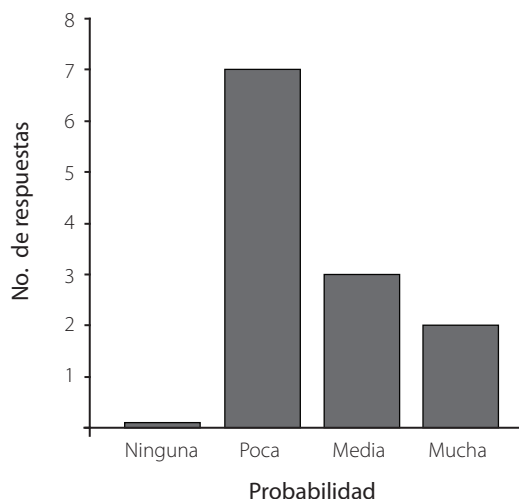


FIG. 1. Respuestas de los entrevistados ante la pregunta: ¿Cuál es la probabilidad de que se desintegren los grupos de investigación con el paso del tiempo?

blecer los patrones también ayudaron a modelar el perfil y características de las entidades modeladas en el SM.

El sistema multiagente

El SM implementado es un resultado concreto de la investigación y toma en cuenta las características establecidas para las entidades que modelan a los investigadores. Además incluye en su funcionamiento la posibilidad de establecer el valor de las variables o parámetros que influyen en el modelo.

Así, en el inicio de la simulación se designan:

- La cantidad de agentes a participar.
- La cantidad de recursos disponibles.
- La probabilidad de establecer vínculos con colegas conocidos.

Seguidamente de manera automática se asignan aleatoriamente según una distribución uniforme las características de los agentes.

La simulación puede ejecutarse en un modo interactivo (Fig. 2), donde es posible visualizar la formación de los grupos en la simulación, además de observar un histograma del tamaño promedio de los grupos, una gráfica del consumo de recursos y el desarrollo del coeficiente de agrupamiento. Al finalizar, este modo produce un informe con el detalle de los resultados de la simulación.

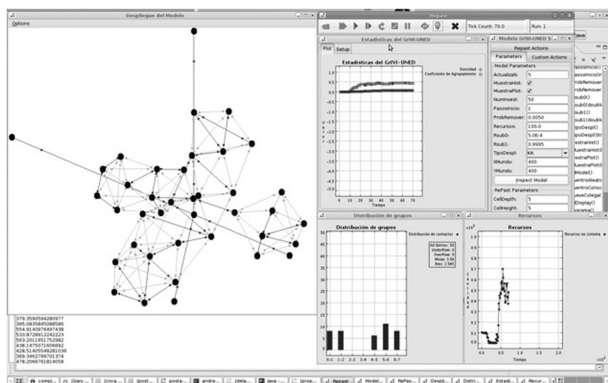


FIG. 2. Visualización en modo interactivo del sistema multiagente con 50 investigadores.

También, es posible llevar a cabo un conjunto de simulaciones en lote (sin visualización), con el fin de obtener más rápidamente un volumen mayor de resultados. De esta forma se facilita el análisis al investigador.

Es necesario aclarar que en cada iteración de la simulación se escoge un subconjunto aleatorio del total de la población a partir de una probabilidad R_0 dada de que surjan encuentros entre investigadores conocidos previamente. Dicho conjunto establece vínculos y forma grupos dependiendo de: la cantidad de recursos disponibles en ese momento de la simulación y el número de colegas con los que cada investigador ha establecido contacto. Para evitar un comportamiento sin plausibilidad social, existe una probabilidad $R_1 = 1 - R_0$ que permite establecer vínculos con colegas desconocidos. Igualmente el modelo considera la posibilidad de separación de algún investigador de un grupo en cada iteración.

En lo que respecta a los recursos disponibles, es de vital importancia aclarar que estos no poseen un valor estático, por el contrario, varían según las entidades del modelo formen grupos, específicamente, al establecer vínculos los recursos se modifican de manera proporcional al costo y el beneficio relativos que implican los investigadores que los establecen.

Finalmente, en cada iteración se calcula el coeficiente de agrupamiento, con el fin de determinar en dicho instante el grado en el que tienden a formarse grupos en la simulación.

Análisis de sensibilidad

Para el análisis de sensibilidad se efectuaron 10 simulaciones con 300 iteraciones cada una, para estimar el grado de estabilidad del MC. Se calcularon la varianza del coeficiente

de agrupamiento, la cual fue de 0,01 y su desviación estándar, cuyo valor fue de 0,07. Dichos resultados indican que las simulaciones poseen un comportamiento estable.

Resultados a nivel macro

El comportamiento de múltiples simulaciones fue revisado según el criterio experto de la vicerrectora de la VI de la UNED.

El SM mostró el comportamiento esperado, a saber, la formación de grupos de investigación dependiendo de la disponibilidad de recursos y de las probabilidades asignadas para los encuentros entre investigadores. No se detectaron situaciones anómalas en ningún momento.

Resultados a nivel micro

Los resultados obtenidos durante las simulaciones se analizaron mediante métodos estadísticos, a saber, se calcularon promedios, varianzas y desviaciones estándar de los coeficientes de agrupamiento resultantes. También se crearon gráficas que muestran el comportamiento de las variables a través de las simulaciones, específicamente, de los recursos, la cantidad de agentes y su relación con respecto al coeficiente de agrupamiento.

En el primer escenario de simulación, los datos indican que el promedio del coeficiente de agrupamiento de las simulaciones es de 0,6 con una varianza de 0,06 y una desviación estándar de 0,08, es decir, existe una tendencia de formación de grupos en el modelo en todas las simulaciones, sin embargo, es importante mencionar que según el modelo no hay un patrón observable que indique algún tipo de dependencia entre el coeficiente de agrupamiento y la cantidad de recursos disponible (Fig. 3). En este caso específico, el aumento de recursos no es sinónimo de aumento en el coeficiente de agrupamiento.

Por otra parte, en el segundo escenario se obtuvo un coeficiente de agrupamiento promedio de 0,45, una varianza de 0,06 y una desviación estándar de 0,07. Los resultados señalan un patrón que relaciona el coeficiente de agrupamiento con la cantidad de investigadores. Conforme la segunda aumenta el coeficiente tiende a disminuir (Fig. 4). Es importante mencionar que la cantidad inicial de investigadores era de 50 y se alcanza un total de 1000 al finalizar las simulaciones, lo cual genera un volumen de datos considerable para el análisis. Cabe aclarar que en todas las instancias del segundo experimento la cantidad de recursos iniciales fue la misma.

Es importante recalcar que en ambos escenarios las normas para el establecimiento de vínculos entre agentes permaneció intacta, es decir, los investigadores dan preferencia a otros colegas conocidos previamente para

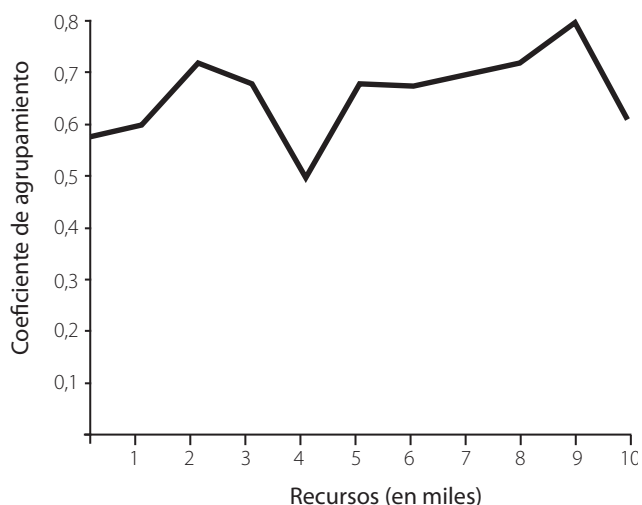


FIG. 3. Variación del coeficiente de agrupamiento con respecto a la cantidad disponible de recursos.

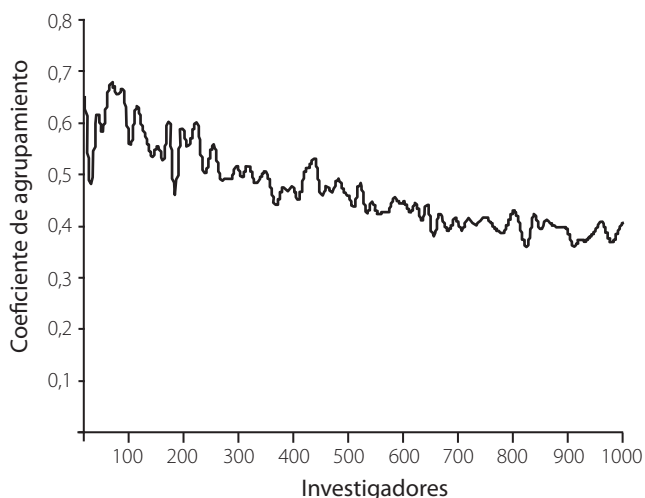


FIG. 4. Variación del coeficiente de agrupamiento con respecto a la cantidad de investigadores.

crear grupos y el tamaño de los grupos varía entre 3 y 7 integrantes.

DISCUSIÓN

Los resultados obtenidos en el primer escenario, es

decir, que los recursos iniciales en el sistema no permiten prever el valor del coeficiente de agrupamiento, indican que el modelo debe afinarse. La presente investigación considera que es probable que deba refinarse el concepto de recurso, particularmente incluir no solamente aspectos económicos sino también infraestructurales y de equipamiento. Es posible que así se logre discriminar con mayor claridad el efecto de la disponibilidad de recursos en la formación de grupos de la UNED. Además es probable que sea necesario recolectar un mayor volumen de datos por un período relativamente largo para afinar el modelo en este sentido.

Por otro lado, el segundo escenario muestra un patrón de comportamiento interesante. Contrario a lo que podría pensarse, el hecho de aumentar la cantidad de participantes en la simulación no genera un mayor nivel de agrupamiento. El SM indica consistentemente a través de las simulaciones que el coeficiente de agrupamiento disminuye conforme aumenta la cantidad de investigadores en el modelo, lo cual se interpreta como una reacción del sistema para compensar el costo y el beneficio relativos de mantener a tantos investigadores. Esto es importante para la VI, ya que podría indicar que los esfuerzos no deben enfocarse necesariamente en un aumento en la cantidad de investigadores de la institución, sino que es necesario previamente establecer las condiciones y espacios que permitan la sostenibilidad de grupos de investigación, con el fin de que estos brinden una mayor cantidad de beneficios a la institución y posteriormente puedan acoger nuevos investigadores para desarrollar nuevas iniciativas.

Las regularidades encontradas no son observables a simple vista en la realidad, lo cual brinda gran valor a las simulaciones, ya que no solo ayudan a comprender el fenómeno social, sino que además permiten formular nuevas teorías para su explicación (Gilbert 2008). El MC creado puede plantearse como una teoría del comportamiento de los grupos de investigación de la UNED. Como teoría, esta debe aproximarse, en un proceso constante, lo más posible al fenómeno observado, mediante métodos cuantificables y verificables empíricamente.

Finalmente es importante rescatar la utilidad del MC llevado a cabo. Esta herramienta facilita y apoya la toma de decisiones en el ámbito de la gestión de la VI, dado que brinda la posibilidad de explorar posibles condiciones para la formación de grupos de investigación en la UNED, mediante un SM basado en datos reales de la institución.

En este sentido es fundamental recalcar que en la medida en que se continúe con una recolección sistemática de datos en la VI, será posible refinar el MC e incluir una mayor cantidad de variables que sirvan como parámetros a las futuras simulaciones. La presente investigación no

es un producto acabado, sino por el contrario, abre una gama de alternativas de desarrollo de mejores modelos, no solo del fenómeno de agrupamiento entre investigadores, sino de cualquier otro susceptible de modelación multiagente, por ejemplo la gestión de recursos de la VI o las relaciones interinstitucionales representan posibles campos de análisis futuros.

AGRADECIMIENTOS

Agradezco a Katya Calderón, cuya visión, motivación, apoyo y acompañamiento hicieron posible esta investigación. Igualmente agradezco a Alvaro de la Ossa por sus valiosos consejos y oportunas observaciones durante el desarrollo del modelo computacional. También agradezco la disponibilidad y amabilidad del equipo de trabajo de la Vicerrectoría de Investigación y de todos los investigadores que participaron de las entrevistas.

REFERENCIAS

- Burger, M.J. 2009. Social context and network formation: An experimental study, *Social Networks* 31: 63-75.
- Carley, K. M. & L. Gasser. 1999, *Computational Organization Theory* en Weiss, G. (Ed.), *Multiagent Systems: a Modern Approach to Distributed Artificial Intelligence*, The MIT Press, Cambridge, Massachusetts, Estados Unidos.
- Fortino, G., A. Garro & W. Russo. 2006. From modeling to enactment of distributed workflows: an agent-based approach, *Proceedings of the 21st ACM Symposium on Applied Computing SAC'06*: 128 (<http://doi.acm.org/10.1145/1141277.1141306>; consultado 10 de junio 2009).
- Gilbert, N. 2008. *Agent-based Models*, ser. *Quantitative Applications in the Social Sciences* 153, SAGE Publications, Londres, Inglaterra.
- Gilbert, N. & K.G. Troitzsch. 2005. *Simulation for the Social Scientist*, Open University McGraw Hill Education, Berkshire, Inglaterra.
- Hudlicka, E. & G. Zacharias. 2004. Approaches for modeling individuals within organizational simulations, *Proceedings of the 36th Winter Simulation Conference WSC'04*: 903-911 (<http://portal.acm.org/citation.cfm?id=1161901>; consultado 21 de agosto 2009).
- Knoke, D. & S. Yang. 2008. *Social Network Analysis*, ser. *Quantitative Applications in the Social Sciences* 154, SAGE Publications, Londres, Inglaterra.
- Lees, M., B. Logan & G. Theodoropoulos. 2007. Distributed Simulations of Agent-Based Systems with HLA, *ACM Transactions on Modeling and Computation* 17.
- Louie, M.A. & K.M. Carley. 2008. Balancing the criticisms: Validating multi-agent models of social systems, *Simulation Modelling Practice and Theory* 16: 242-256.
- Oprea, M. 2004. *Applications of Multi-Agent Systems*, IFIP International Federation for Information Processing: Information Technology, Springer, Boston, Estados Unidos.
- REPAST. 2009. Recursive Porous Agent Simulation Toolkit (<http://repast.sourceforge.net/>; consultado 1 de marzo 2009)
- Troitzsch, K. 2009. Perspectives and challenges of agent-based simulation as a tool for economics and other social sciences, *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems AAMAS 2009*: 35-42, International Foundation for Autonomous Agents and Multiagent Systems, Richland, Carolina del Sur, Estados Unidos.
- Watts, D.J. & S. Strogatz. 1998. Collective dynamics of "small world" networks, *Nature* 393.

