

Modelos estocásticos orientados en el actor utilizando RSiena (V). Estimación de un modelo para la evolución de redes sociales.

Alvaro Uzaheta¹ y Alejandro Espinosa-Rada¹

¹ETH Zürich, Social Networks Lab

RESUMEN

Este texto es el quinto guión de una serie de seis guiones escritos en el lenguaje y ambiente de programación R que conjuntamente constituyen la introducción al software RSiena para estimar modelos estocásticos orientados en el actor para redes dinámicas. Este quinto guión presenta los pasos para estimar los parámetros de un modelo previamente especificado en RSiena. El texto está acompañado por el código en R para reproducir y estimar los modelos.

Palabras clave: *Análisis de redes Sociales - Modelos estocásticos orientados en el actor - Análisis estadísticos de redes sociales - RSiena - R.*

ABSTRACT

This article is the fifth document of six scripts written in the language and environment R, which introduces the RSiena software to estimate the stochastic actor-oriented models for dynamic networks. This fifth script shows the steps to estimate the parameters of a previously specified model in RSiena. The text incorporates the R code into the document to reproduce and estimate the models.

Key words: *Social Network Analysis - Stochastic actor-oriented model - Statistical analysis for social networks - RSiena - R.*

¹Contacto con los autores: Alejandro Espinosa-Rada (alejandro.espinosa@qess.ethz.ch)

INTRODUCCIÓN

Los materiales para esta serie incluyen seis guiones (*scripts*) para utilizar el paquete **RSiena**. Dicho paquete se encuentra escrito en el entorno y lenguaje de programación **R** (2022) enfocado a la estadística computacional y de gráficos. Los guiones, a su vez, se encuentran disponibles en la página oficial¹ del programa para el análisis estadístico de datos de redes SIENA (acrónimo de *Simulation Investigation for Empirical Network Analysis*), con énfasis en redes sociales, el cual cuenta con un manual de usuarios (Ripley et al., 2022).

Este guión es la quinta contribución de una serie de seis guiones que introducen el software **RSiena**. Este guión brinda información adicional sobre la definición de los datos como variables en **RSiena** y de como especificar modelos para su posterior estimación. El primer guión de esta serie (Espinosa-Rada, 2022a) entrega un panorama general de cómo se estima un modelo SAOM [acrónimo para *stochastic actor-oriented model*] (mayor detalle sobre los modelos en: Snijders, 2001, 2017 y en castellano ver Espinosa-Rada, 2023). El segundo guión, en cambio, introduce el ambiente **R** y los distintos tipos de estructuras de datos que pueden ser utilizados para importar o exportar los datos para los análisis (Espinosa-Rada, 2022b). El tercer guión, en cambio, presenta un ejemplo de análisis descriptivo a realizar antes de la estimación de un modelo SAOM (Espinosa-Rada y Uzaheta, 2022). El cuarto guión amplía la información sobre la definición de datos como variables para **RSiena** y especificación de modelos introducida en el primer guión (Uzaheta y Espinosa-Rada, 2022).

A continuación, se presenta en el texto los códigos a ejecutar en el programa **R** anteceditos por una breve explicación. En color “()” se encuentran las líneas de códigos o comandos a ejecutar y/o que pertenecen a funciones propias del programa. De la misma forma, líneas de códigos anteceditos por # corresponden a comentarios y no comandos. Finalmente, líneas anteceditas por > corresponden a la salida que se imprime en

consola del código presentado en el mismo bloque.

Estudio sobre Amigos Adolescentes y Estilos de Vida

En este tutorial utilizaremos el estudio sobre Amigos Adolescentes y Estilos de Vida, estudio realizado en 1995. Dicha investigación analizó a una cohorte de estudiantes en una escuela en el oeste de Escocia. La investigación registró información sobre tres años comenzando en 1995, cuando los/las alumnos/as tenían en promedio 13 años, y terminó en 1997. De los/las 160 alumnos/as, 129 de ellos/as estuvieron presente en las tres medidas. Para la red de amistad investigada, a cada uno de los estudiantes se les pidió que registraran un máximo de 12 otros estudiantes. De la misma forma, a los estudiantes se les pidió que registraran algunas variables de comportamiento tales como estilos de vida, realización de deporte, consumo de tabaco, alcohol y consumo de cannabis². Los datos se encuentran disponibles en el paquete **RSiena** y se pueden cargar al ambiente de trabajo en **R** usando los siguientes comandos.

```
library(RSiena)
```

```
friend.data.w1 <- s501
friend.data.w2 <- s502
friend.data.w3 <- s503
drink <- s50a
smoke <- s50s
```

Transformando los datos en un objeto de tipo *siena* y definiendo la especificación del modelo

El proceso de estimación en **RSiena** espera un objeto de tipo “*siena*” que contiene la variable dependiente en el modelo y las covariantes usadas en la especificación del modelo. Una descripción extensa de cada una de las funciones usadas para definir este tipo de objetos y las diferentes variantes disponibles se pueden encontrar en previas entregas (Uzaheta y Espinosa-Rada, 2022). Los

¹ Página oficial del software **RSiena**: <https://www.stats.ox.ac.uk/~snijders/siena/>

² Para más información sobre la base de datos, codificación de las variables y estudios aparejados, visitar:

https://www.stats.ox.ac.uk/~snijders/siena/s50_data.htm. Los datos también pueden descargarse en: http://www.stats.ox.ac.uk/~snijders/siena/s50_data.zip

siguientes comandos crean el objeto con los datos que se usarán para estimar los modelos.

```
friendship <- sienaDependent(
  array(c(friend.data.w1, friend.data.w2,
    friend.data.w3),
    dim = c(50, 50, 3)))
smoke1 <- coCovar(smoke[, 1 ])
alcohol <- varCovar(drink)
mydata <- sienaDataCreate(friendship, smoke1,
  alcohol)
```

Ahora que tenemos un objeto del tipo "siena" podemos crear a continuación un objeto con los efectos que definen la especificación del modelo:

```
myeff <- getEffects(mydata)
myeff <- includeEffects(myeff, transTrip, cycle3)
myeff <- includeEffects(myeff, egoX, altX, simX,
  interaction1 = "alcohol")
myeff <- includeEffects(myeff, simX,
  interaction1 = "smoke1")
myeff
```

El modelo contiene una especificación básica de efectos endógenos de la red como densidad, reciprocidad, transitividad y reciprocidad indirecta (o ciclos). Estos efectos son parte de los efectos estructurales importantes descritos en el manual de **RSiena** (Ripley et al., 2022, Sección 5.2). Adicionalmente, se incluyen efectos usando las variables covariantes monádicas (tratadas a continuación como constantes): consumo de alcohol y consumo de cigarrillo. Para ambas se incluye un efecto modelando la influencia en la formación de los vínculos de amistad por la similitud en el consumo de alcohol o cigarrillos (procesos de selección por homofilia). Para la variable consumo de alcohol, también se agregan los efectos ego y de los alters (alias *alteris*)³ para estudiar los cambios en la formación de los vínculos de amistad dependiendo de que tan frecuente consumen alcohol.

Especificaciones de la estimación

El último elemento que se requieren definir son las especificaciones que se utilizan durante la estimación. La función `sienaAlgorithmCreate()` crea un objeto con estas especificaciones. En general se recomienda usar las especificaciones que vienen por defecto. Aquí solo mencionaremos dos argumentos. No

obstante, para ver todas las especificaciones que se pueden cambiar y sus posibles valores use la documentación de esta función (`?sienaAlgorithmCreate`). El primer argumento que mencionaremos es `projname`, este permite definir el nombre del archivo de registro que se creará en la misma carpeta donde se fijó el directorio de trabajo. El segundo argumento es `seed` que permite fijar la semilla usada para la generación de valores aleatorios usados durante la estimación. Fijar la semilla permite reproducir exactamente los mismos resultados cuando se realizan diferentes corridas del modelo.

```
myalgorithm <- sienaAlgorithmCreate(
  projname = 's50_3')
```

Estimando el modelo

Una vez todos los elementos requeridos para la estimación se han definido se procede a correr el algoritmo de estimación de los parámetros. La función `siena07()` realiza la estimación. Cuando se ejecuta este comando, se abre una interfaz gráfica de usuario con una imagen de la ciudad de Siena en Italia e información sobre el progreso de la estimación.

```
ans <- siena07(myalgorithm, data = mydata,
  effects = myeff)
```

En el caso que abrir la interfaz gráfica de usuario genere dificultades o no sea necesario tener información del progreso del proceso de estimación, se puede fijar el argumento `batch` a valor verdadero `batch = TRUE`. Información reducida sobre el progreso de la estimación será impresa en la consola de **R** en este caso. Si se fija el argumento `verbose` a valor verdadero `verbose = TRUE`, se imprimirá en la consola una información detallada del proceso de estimación la cual será añadida al archivo creado en el paso anterior. Para usar múltiples procesadores durante la estimación, es necesario fijar los argumentos `nbrNodes = 2`, `useCluster = TRUE`. En este caso se usarían dos procesadores, ajuste este parámetro al valor deseado, pero recuerde dejar un procesador libre para que haga las veces de orquestador y maestro de los otros. Configuraciones más avanzadas del procesamiento en paralelo son posibles usando el argumento `clusterString`,

³ Se usan los términos del latín para referirse a la persona que nomina y al nominado, respectivamente en este contexto.

mire la documentación de `siena07()` y la del paquete `parallel`.

Para funcionalidades avanzadas como verificar la bondad de ajuste del modelo (Lospinoso y Snijders, 2019) es recomendable conservar las redes simuladas en la última etapa de estimación del algoritmo. Para ello fije el argumento `returnDeps = TRUE`. La función `siena07()` crea un objeto de tipo `sienaFit` en todos los casos, y cuando `returnDeps = TRUE` este objeto tendrá un componente adicional llamado `sims` con las redes simuladas en estructura de lista anidada con las redes en forma de lista de vínculos.

La salida del proceso de estimación

El archivo `s50_3.txt` en la carpeta del directorio de trabajo contiene los resultados del proceso de estimación. Una versión resumida se puede obtener en la consola ejecutando `ans` y una versión detallada con `summary(ans)`. Dependiendo de la semilla para la aleatorización usada y la especificación del modelo los resultados se parecen a la siguiente salida:

```
ans
>Estimates, standard errors and convergence t-ratios
>
>
>                Estimate Standard Convergence
>                Error      t-ratio
>
>Rate parameters:
> 0.1 Rate parameter period 1  6.5873  (1.1455)
> 0.2 Rate parameter period 2  5.2975  (0.9301)
>
>Other parameters:
> 1. eval outdegree (density) -2.7343  (0.1409) -0.0944
> 2. eval reciprocity         2.4380  (0.2350) -0.0746
> 3. eval transitive triplets  0.6591  (0.1607) -0.0619
> 4. eval 3-cycles            -0.0783  (0.3135) -0.0462
> 5. eval smoke1 similarity    0.1975  (0.2082) -0.0021
> 6. eval alcohol alter       -0.0140  (0.0738) -0.0372
> 7. eval alcohol ego         0.0511  (0.0784) -0.0414
> 8. eval alcohol similarity   0.7300  (0.2979) -0.0841
>
>Overall maximum convergence ratio:  0.1706
>
>
>Total of 2438 iteration steps.
```

Los valores reportados como "Convergence t-ratio" corresponde a las razones t usadas para verificar la convergencia del proceso de estimación. Note que esto no corresponde a las estadísticas t para el juzgamiento de la hipótesis de significancia de los parámetros. La sección 6.2 del manual brinda información detallada sobre esto (Ripley et al., 2022). La regla empírica para evaluar convergencia sugiere que todas las razones t deben ser idealmente menores a 0.1 en valor absoluto, y que la máxima razón de convergencia general ("Overall maximum convergence ratio") sea menor que 0.25. Esto significa una buena convergencia del algoritmo de estimación. Este es el caso para los resultados presentados para el modelo estimado. Si no fuese el caso, lo mejor es continuar con el proceso de estimación usando como punto de partida los

resultados obtenidos hasta el momento. Para ello utilice el siguiente comando:

```
ans <- siena07(myalgorithm, data = mydata,
              effects = myeff, prevAns = ans)
```

Al usar este comando la fase 1 será omitida utilizando en cambio la información contenida en los resultados previos. Cabe resaltar que esto solo es posible si se quiere continuar el proceso de estimación y la especificación del modelo no cambia. Es importante revisar que las estimaciones contenidas en `ans` sean razonables, en caso contrario es mejor realizar el proceso de estimación de nuevo. Las secciones 6.3 y 6.5 del manual brindan mayor información sobre los valores iniciales usados en el algoritmo y información adicional en el caso que el modelo presenta problemas de convergencia (Ripley et al., 2022).

Debido a que el algoritmo de estimación es basado en simulaciones aleatorizadas de la evolución de la red, siempre se presentarán pequeñas diferencias entre diferentes corridas del algoritmo de estimación. Se recomienda fijar el número de iteraciones en fase 3 a un valor mayor que el se tiene por defecto, por ejemplo 5000. Para ello fije el argumento `n3` en la invocación de `sienaAlgorithmCreate()` a el valor deseado `n3 = 5000`.

El objeto `ans` contiene todos los resultados de la estimación. Por ejemplo, `ans$theta` contiene el vector con las estimaciones de los parámetros, `ans$se` contiene un vector con los errores estándar de las estimaciones y `ans$scovtheta` contiene la matriz de varianzas y covarianzas de las estimaciones.

Adicionalmente, **RSiena** incluye métodos para poder obtener los resultados en versiones listas para publicación o que sean más amigables visualmente. El comando `siena.table(ans)` producirá un archivo en la carpeta del directorio de trabajo con una tabla en formato LaTeX lista para ser incluida en un documento, haciendo uso del paquete `xtable` se puede modificar esta salida programáticamente. Usando `siena.table(ans, type = "html")` se producirá una tabla en formato html que podrá ser incluida en un documento de Word o copiada/abierto en Excel para continuar la edición de la tabla. Por último, el comando `?print.sienaFit` permite ver información adicional de las posibilidades de edición.

Pruebas estadísticas para el juzgamiento de hipótesis

Existen tres tipos de pruebas estadísticas para el juzgamiento de hipótesis que se pueden

realizar con **RSiena**. El primero corresponde a las pruebas tipo t-test para parámetros individuales que se puede hacer calculando la razón entre la estimación del parámetro y su error estándar. Bajo la hipótesis nula que el parámetro es 0, este estadístico tiene asintóticamente una distribución normal estándar.

Un segundo tipo de pruebas corresponden a la *score-type* para parámetros individuales o para múltiples parámetros al mismo tiempo. Los parámetros pueden ser restringidos y declarados para ser usados para el test usando los argumentos `fix` and `test` en la función `setEffects()`. Por ejemplo, para hacer el test para el efecto de la raíz cuadrada de la popularidad en indegree se usa el siguiente comando.

```
myeff <- setEffects(myeff, inPopSqrt, fix = TRUE,
  test = TRUE, initialValue = 0.0)
ans <- siena07(myalgorithm, data = mydata,
  effects = myeff)
```

La prueba se lleva a cabo usando el comando `score.Test(ans)`. Ver la documentación de esta función para obtener mayor información al respecto. El tercer tipo de prueba son los Wald-test. Para parámetros individuales y múltiples se obtienen usando las funciones `Wald.RSiena()` y `Multipar.RSiena()`, ver la documentación y el manual para su uso.

Una aplicación de la prueba de tipo score que requiere un tratamiento especial es cuando se hace el test de heterogeneidad en el tiempo de los parámetros, es decir, los parámetros cambian de valor para diferentes períodos. El test se puede obtener ejecutando el comando `sienaTimeTest()` y es presentado en la literatura en Lospinoso et al. (2011).

Comentarios finales

En el siguiente guión se revisó cómo estimar un modelo usando el software **RSiena**. Brindando especial atención a las especificaciones del algoritmo de estimación y como continuar el proceso de estimación en el caso que no se presente convergencia para uno o más parámetros. Adicionalmente, se revisó las diferentes pruebas estadísticas para el juzgamiento de hipótesis para los parámetros del modelo disponibles en **RSiena**. Se menciona los diferentes métodos disponibles para llevar los resultados del modelo a tablas listas para publicación.

El siguiente guión de esta serie (Espinosa-Rada y Uzaheta, 2023) se concentrará principalmente en la estimación de modelos de coevolución de redes y comportamiento.

Esperamos que esta introducción permita dar una aproximación a los modelos estocásticos basados en el actor a través del uso práctico del programa, acorde al estado actual de su desarrollo y que sea accesible a la comunidad que habla español.

Agradecimientos

Agradecemos particularmente a Tom A.B. Snijders (University of Oxford y University of Groningen) por su amabilidad en permitirnos difundir, modificar y extender los siguientes guiones los cuales contaron también con contribuciones previas de Robin Gauthier, Ruth Ripley, Johan Koskinen, Paulina Preciado, Zsofia Boda y Christian Steglich. En un principio estos documentos iban ser una traducción textual a los guiones disponibles en la página de RSiena. No obstante, por sugerencia de Tom A. B. Snijders los guiones traducidos y modificados adquieren con esta contribución independencia de sus homólogos en inglés con el objeto de facilitar su difusión, mantención y responsabilidad en su contenido en español del cual esperamos reflejar adecuadamente sus intenciones originales.

REFERENCIAS

Espinosa-Rada, Alejandro (2022a). Modelos estocásticos orientados en el actor utilizando RSiena (I): Guión básico introductorio. *Redes: revista hispana para el análisis de redes sociales*, 33(1), 0092-99. DOI: <https://doi.org/10.5565/rev/redes.936>

Espinosa-Rada, Alejandro (2022b). Modelos estocásticos orientados en el actor utilizando RSiena (II): Formato de los Datos. *Redes. Revista hispana para el análisis de redes sociales*, 33(1), 100-111. DOI: <https://doi.org/10.5565/rev/redes.937>

Espinosa-Rada, Alejandro (2023). "Usos contemporáneos de los modelos estocásticos orientados en el actor (SAOMs) para redes dinámicas". En: Francisca Ortiz y Alejandro Espinosa-Rada (Eds.). *Redes Sociales: teoría, métodos y aplicaciones en América Latina*. CIS-Madrid. *En prensa*.

Espinosa-Rada, Alejandro, y Uzaheta, Alvaro (2023). Modelos estocásticos orientados en el actor utilizando RSiena (VI). Coevolución de redes sociales y comportamientos individuales. *Redes. Revista Hispana para el Análisis de Redes Sociales*, 34(1), 126-132.

Espinosa-Rada, Alejandro, y Uzaheta, Alvaro (2023). Modelos estocásticos orientados en el actor utilizando RSiena (III): Análisis Descriptivo. *Redes. Revista hispana para el análisis de redes sociales*, 33(2), 196-

202. DOI:
<https://doi.org/10.5565/rev/redes.956>

Lospinoso, Joshua A., Schweinberger, Michael, Snijders, Tom A. B., & Ripley, Ruth M. (2011). Assessing and accounting for time heterogeneity in stochastic actor oriented models. *Advances in data analysis and classification*, 5(2), 147-176.

Lospinoso, Joshua, y Snijders, Tom A. B. (2019). Goodness of fit for stochastic actor-oriented models. *Methodological Innovations*, 12(3): 1-18.

DOI:
<https://doi.org/10.1177/2059799119884282>

Ruth Ripley, Tom A.B. Snijders, Zsófia Boda, Andras Vörös, and Paulina Preciado (2022). *Manual for SIENA version 4.0*. Oxford: University of Oxford, Department of Statistics, <http://www.stats.ox.ac.uk/~snijders/siena/>.

Snijders, Tom A. B. (2001) The statistical evaluation of social network dynamics, *Sociological Methodology* - 2001, 40: 361-395.

Snijders, Tom A. B. (2017). [Stochastic Actor-Oriented Models for Network Dynamics](https://doi.org/10.1146/annurev-statistics-060116-054035). *Annual Review of Statistics and Its Application*, 4, 343-363. DOI: <http://dx.doi.org/10.1146/annurev-statistics-060116-054035>
[Here is the e-print access to this article.](#)

Uzaheta, Alvaro y Espinosa-Rada, Alejandro (2022). Modelos estocásticos orientados en el actor utilizando RSiena (IV): Formato de las variables. *Redes: revista hispana para el análisis de redes sociales*, 33(2), 0203-209. DOI: <https://doi.org/10.5565/rev/redes.957>

Remitido: 04-11-2022

Corregido: 05-11-2022

Aceptado: 06-11-2022



© Los autores