Advanced Quantitative Methods: Networks and spatial data

Johan A. Elkink

University College Dublin

4 May 2018
1 Introduction

2 Spatial autocorrelation

3 Spatial Autoregressive Models

4 Exponential Random Graph Models

5 Stochastic Actor-Oriented Models
Outline

1. Introduction
2. Spatial autocorrelation
3. Spatial Autoregressive Models
4. Exponential Random Graph Models
5. Stochastic Actor-Oriented Models
Dyads

A dyad is a pair of actors in the network, often with characteristics associated with the two nodes, and variables associated with their ties.

**Dyadic data** is very common in international relations and usually studied as panel data, but with clustered standard errors.

<table>
<thead>
<tr>
<th>ctry1</th>
<th>ctry2</th>
<th>year</th>
<th>trade</th>
<th>war</th>
<th>dem1</th>
<th>dem2</th>
</tr>
</thead>
<tbody>
<tr>
<td>NL</td>
<td>US</td>
<td>2000</td>
<td>49.2</td>
<td>0</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>NL</td>
<td>US</td>
<td>2001</td>
<td>50.3</td>
<td>0</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IE</td>
<td>RU</td>
<td>1974</td>
<td>20.3</td>
<td>0</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>US</td>
<td>VN</td>
<td>1967</td>
<td>5.5</td>
<td>1</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In political science we are generally working with dyads, but with the introduction of network statistics in the discipline, we are starting to look beyond this. The number of possible configurations quickly expands:

Network terminology

Points are called nodes or vertices.

Lines are called ties, edges, or arcs.

Ties can be directed (asymmetric $W$) or undirected (symmetric $W$).
Interdependence

Observations are often not as independent as we assume.

- Observation at time $t$ affected by observation at $t - 1$.
- Observation in unit $i$ affected by neighbour $j$.
- Connection between $i$ and $j$ affected by relation between $j$ and $k$.

((Cingolani, Piccardi and Tajoli, 2015)

When the tie is the dependent variable: network statistics.
When the node is the dependent variable: spatial econometrics.
Outline

1. Introduction

2. Spatial autocorrelation

3. Spatial Autoregressive Models

4. Exponential Random Graph Models

5. Stochastic Actor-Oriented Models
Adjacency matrix

An easy formulation of a network is the **adjacency**, **connection** or **contiguity** matrix.

\[
W = \begin{bmatrix}
0 & 1 & 1 & 0 & 0 & 1 \\
1 & 0 & 1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 1 & 1 \\
0 & 0 & 1 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 
\end{bmatrix}
\]
Note that $\tilde{\mathbf{W}}\mathbf{y}$ is equivalent to calculating a variable to captures the average value of $\mathbf{y}$ across all neighbours.
No spatial autocorrelation
Negative spatial autocorrelation
Positive spatial autocorrelation
“Space”

We can think of space in two or three different ways:

- A **geostatistics perspective** takes each observation at their \((x, y)\) coordinates and models space as a continuous variable.

- An **object view** or **lattice perspective** takes each observation as related to a defined set of neighbours and thus models space as a discrete variable.

- In some instances, a **grid perspective** takes the middle ground between those two (e.g., Friel and Pettitt, 2004).

In the remainder of this lecture, we assume the object view, which means *de facto* that space is modelled as a network—and all of the following applies to any kind of network, not just those geographically defined.

(Anselin, 2002, 255)
Moran’s I

\[ I = \frac{\sum_i \sum_j \tilde{w}_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i \sum_j \tilde{w}_{ij}} \cdot \frac{n}{\sum_i (x_i - \bar{x})^2} \sim N(\mu_I, \sigma_I^2) \]

\[ \mu_I = E(I) = \frac{-1}{n - 1} \]

\[ \sigma_I^2 = Var(I) = \frac{n^2 S_1 - nS_2 + 3S_0^2}{S_0^2(n^2 - 1)} \]

where

\[ S_0 = \sum_i \sum_j (w_{ij} + w_{ji}) \]

\[ S_1 = \frac{1}{2} \sum_i \sum_j (w_{ij} + w_{ji})^2 \]

\[ S_2 = \sum_i \sum_j (\tilde{w}_{ij} + \tilde{w}_{ji})^2 \]
Moran’s $I$ can only be calculated with a known $W$ matrix. Higher order lags are also possible, e.g. $W^2$. 

```r
library(ape)
Moran.I(y, W)
Moran.I(residuals(lm(y ~ x1 + x2)), W)
```
Example: democracy (Elkink, 2011)

Spatial clustering, Polity IV dichotomized, $I(y_t)$ and $I(\Delta y_t)$
Outline

1. Introduction

2. Spatial autocorrelation

3. Spatial Autoregressive Models

4. Exponential Random Graph Models

5. Stochastic Actor-Oriented Models
Spatial econometrics: motivation

From a **theoretical viewpoint**: Spatial clustering can occur due to regional competition, spillover effects, regional substitution effects.

From a **statistical viewpoint**: Spatial dependence can arise from unobservable latent variables that are spatially correlated or as a result of the arbitrariness of the boundaries between spatial units (the **modifiable areable unit problem**).

(Fotheringham and Wong 1991, Brueckner 2003, LeSage

Spatial processes

Spatial autocorrelation has processes somewhat analogous to serial autocorrelation.

Spatial error process: \( y = X\beta + u, \quad u = \lambda Su + \varepsilon. \)

Spatial lag process: \( y = \rho Wy + X\beta + \varepsilon. \)

Much less common are:

Spatial moving average process: \( y = X\beta + u, \quad u = \lambda S\varepsilon + \varepsilon. \)

Spatial error components: \( y = X\beta + u, \quad u = \lambda S\eta + \varepsilon \) (with uncorrelated \( \eta \) and \( \varepsilon \)).

\( W \) and \( S \) are adjacency matrices, \( \rho \) and \( \lambda \) spatial parameters.

(Anselin, 1988, 2002; Darmofal, 2015, 40–41)
Ignoring spatial autocorrelation

If we ignore the spatial effect, the impact depends on the type of spatial process:

Spatial error: similar to time-series, OLS estimates coefficient will be unbiased, but the variance-covariance matrix will be biased, typically downwards.

Spatial lag: OLS estimates will be biased and, unlike in time-series, inconsistent.

(Darmofal, 2015, 32–34)
Checking residuals

\[ I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{e'W e}{e'e} \sim N(\mu_I, \sigma^2_I) \]

\[ LM_{\text{err}} = \frac{n^2 \left( \frac{e'W e}{e'e} \right)^2}{\text{tr}(W'W + W^2)} \sim \chi^2(1) \]

\[ LM_{\text{lag}} = \frac{n^2 \left( \frac{e'W y}{e'e} \right)^2}{(WX\hat{\beta}^{OLS})'MWX\hat{\beta}^{OLS} / \sigma^2 + \text{tr}(W'W + W^2)} \sim \chi^2(1), \]

with \( LM_{\text{err}} \) and \( LM_{\text{lag}} \) referring to tests for spatial error and spatial lag processes, respectively.

(Anselin and Hudak, 1992, 502)
Spatial multiplier

We can somewhat rearrange to obtain:

Spatial error process: \( y = X\beta + (I - \lambda S)^{-1}\varepsilon. \)

Spatial lag process: \( y = (I - \rho W)^{-1}(X\beta + \varepsilon). \)

The matrix \((I - \rho W)^{-1}\) is called a **spatial multiplier** (in economic input-output modelling this is called a **Leontief inverse**).

(Franzese and Hays, 2007; Darmofal, 2015, 33)
Interpretation

The spatial multiplier implies that a change in $x_i$ will have a direct impact on $y_i$, but also an indirect, spillover effect on all other observations:

$$\frac{\partial y}{\partial X} = \beta' \otimes (I - \rho W)^{-1},$$

where $\otimes$ is a Kronecker product. We therefore obtain a $n \times nk$ matrix, with a derivative for each observation towards each $x$, each $\beta$—all changes in $x$ affect all $y$ (unless some derivatives are zero).

We can split this in $k$ separate matrices $\beta_k(I - \rho W)^{-1}$, which in each cell provides the long-run impact of a change in $x_{jk}$ on $y_i$. On the diagonal are both direct and indirect (feedback loop) effects of $x_{ik}$ on $y_i$.

(Gentle 2007, 155; Darmofal 2015, 107; see also LeSage and Pace 2009, 33–41)
Ordinary Least Squares

OLS estimation of $\rho$ leads to a biased and inconsistent estimate:

$$E(\hat{\rho}^{OLS}) = \rho + (y'W'Wy)^{-1}y'W'\varepsilon.$$

Unlike in linear regression, where we assume $E(X'\varepsilon) = 0$, this assumption cannot reasonably be made for $E(y'W'\varepsilon)$.

This bias does not vanish as the sample size increases.

(Darmofal, 2015, 32–33)
Maximum Likelihood

We can estimate using the following loglikelihoods:

Spatial error process (SEM):
\[
\ell = -\frac{1}{2} \log(|\Omega|) - \frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\sigma_{\varepsilon}^2) - \frac{1}{2\sigma_{\varepsilon}^2} (y - X\beta)'\Omega^{-1}(y - X\beta).
\]

Spatial lag process (SAR):
\[
\ell = \log(|A|) - \frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\sigma_{\varepsilon}^2) - \frac{1}{2\sigma_{\varepsilon}^2} (y - \rho Wy - X\beta)'(y - \rho Wy - X\beta),
\]
where $A = (I - \rho W)$ and $\Omega = [(I - \lambda S)'(I - \lambda S)]^{-1}$.

(Darmofal, 2015, 97–103)
Computation

The computationally intensive part of the estimation is 
\[ \log(|A|) = \log(|I - \rho W|), \]
especially when \( W \) is large, since it needs an update for each value of \( \rho \).

This can be improved using

\[ \log(|A|) \approx \log \left( \prod_j (1 - \rho \omega_j) \right) = \sum_j \log(1 - \rho \omega_j), \]

where \( \omega_j \) are the eigenvalues of \( W \).

(Darmofal, 2015, 98)
Binary Spatial Autoregressive Models

In political science, dependent variables are often binary. Spatial models for binary data exist, and can be understood using the latent variable formulation of the binary model:

$$y = \begin{cases} 1, & y^* > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$y^* = \rho Wy^* + X\beta + u$$

$$u = \lambda Su + \varepsilon.$$

The BSAR model (i.e. $\lambda = 0$) is best estimated using a Gibbs sampler (LeSage, 2000) or, if $N$ is large and $\rho$ not too high, a linearized GMM estimator (Klier and McMillen, 2008).

Similarly, models exist for count, multinomial, and ordinal data with spatial interdependence.

(Calabrese and Elkink, 2014)
Outline

1. Introduction

2. Spatial autocorrelation

3. Spatial Autoregressive Models

4. Exponential Random Graph Models

5. Stochastic Actor-Oriented Models
Exponential Random Graph Model

In an **Exponential Random Graph Model (ERGM)**, the network structure as a whole is the dependent variable.

Independent variables include node- and tie-characteristics, but also summary statistics on the network structure, which we will denote with \( G_k \) for \( K \) different summary statistics (varies by model specification), for example:

**Total number of ties**  
\[ G_k = \sum_i^N \sum_j^N w_{ij} \]

**Reciprocity**  
\[ G_k = \sum_i^N \sum_j^N w_{ij} w_{ji} \]

**Transitivity**  
\[ G_k = \sum_i^N \sum_j^N \sum_k^N w_{ij} w_{jq} w_{iq} \]

(Thomas Grund, *Advanced Social Network Analysis*, ERGM slides,  
http://thomasgrund.weebly.com/teaching.html)
ERGM: node characteristics

The node- and tie-characteristics are entered in a similar fashion.

Matching nodes
\[ G_k = \sum_i^N \sum_j^N w_{ij} I(x_i = x_j) \]

Node factors
\[ G_k = \sum_i^N \sum_j^N w_{ij} x_i \]
\[ G_k = \sum_i^N \sum_j^N w_{ij} x_j \]
\[ G_k = \sum_i^N \sum_j^N w_{ij} x_i x_j \]

Tie factors
\[ G_k = \sum_i^N \sum_j^N w_{ij} z_{ij} \]

\( I(\cdot) \) is an indicator matrix which is 1 if the condition is true, 0 otherwise.

ERGM: model

The dependent variable here is the network itself, $W$, whereby we consider a probability distribution over all $M$ possible networks $W_m$.

\[
Pr(W_m) = \frac{\exp \left( \sum_k^K G_{mk} \theta_k \right)}{\sum_i^M \exp \left( \sum_k^K G_{lk} \theta_k \right)} = \frac{\exp \left( \sum_k^K G_{mk} \theta_k \right)}{c}
\]

By optimizing this probability with respect to $\theta$, we therefore obtain the parameters that make the network we observe most likely, with the parameters weighting the various summary statistics.

(Cranmer and Desmarais 2011; Lusher, Koskinen and Robins 2012, 53)
ERGM: log-odds formulation

Note that typically, the nominator $c$ is intractable.

Consider, we have a matrix $W_{ij}$, which differs from $W_{ij}^*$ only in that $w_{ij} = 1$ and $w_{ij}^* = 0$, with all other connections the same. The log-odds are then:

$$
\log \left( \frac{Pr(W_{ij})}{Pr(W_{ij}^*)} \right) = \log \left( \frac{\exp \left( \sum_{k}^{K} G_{(ij)k} \theta_k \right) / c}{\exp \left( \sum_{k}^{K} G_{*(ij)k} \theta_k \right) / c} \right)
$$

$$
= \log \left( \frac{\exp \left( \sum_{k}^{K} G_{(ij)k} \theta_k \right)}{\exp \left( \sum_{k}^{K} G_{*(ij)k} \theta_k \right)} \right)
$$

$$
= \sum_{k}^{K} G_{(ij)k} \theta_k - \sum_{k}^{K} G_{*(ij)k} \theta_k
$$

$$
= \sum_{k}^{K} \left( G_{(ij)k} - G_{*(ij)k} \right) \theta_k
$$
As we’ve seen before, such log-odds relationship can be modelled using a logistic regression:

\[
\text{logit}(w_{ij} = 1 | W = W^*, \theta) = \sum_{k} \left(G_{(ij)_k} - G^*_{(ij)_k}\right) \theta_k
\]

So the coefficients of the ERGM model, \(\theta\), are the logistic regression coefficients representing the change in the log-odds of a tie based on a change in the summary statistics \(G_m\) if that tie changes.

If we only include tie-characteristics, ignoring network summary statistics, we would replicate a regular logistic regression.

Outline

1. Introduction
2. Spatial autocorrelation
3. Spatial Autoregressive Models
4. Exponential Random Graph Models
5. Stochastic Actor-Oriented Models
Motivation

Spatial autoregressive models explain variation in node attributes based on given network ties and node attribute changes elsewhere in the network.

Exponential random graph models explain variation in ties based on given node attribute and tie changes elsewhere in the network.

But what if node changes affect tie formation, and ties affect node changes?

**Stochastic Actor-Oriented Models (SAOMs)** explicitly model both interdependencies.

SAOMs are **dynamic** models, capturing the **coevolution** of nodes and ties.
SAOM: assumptions

SAOMs require longitudinal data: the same nodes and their ties observed over different points in time.

1. Nodes control outgoing ties.
2. Time is modelled as continuous time—we observe discrete snapshots, but changes can happen at any time.
3. Observed large changes are assumed to be due to a series of small changes; changes take place only one at a time.
4. Changes in ties and node attributes are a function of the previous network, not longer memory lags.
5. The model is stochastic: some changes are random.

SAOM: objective function

Each node \( i \) in the model has two functions on which decisions are based, an **objective function** which determines the “preference” for each potential change and a **rate function** which determines when the actor can change.

The objective function is similar to the ERGM equation, but summary statistics are *from the perspective of the node*:

\[
    f_i = \sum_{k}^{K} G_{ki} \theta_k,
\]

whereby \( G_{ki} \) is (potentially) a function of node attributes \( x \), tie attributes \( z \) and network structure \( W \). The same log-odds approach as ERGM applies, for any change:

\[
    \log \left( \frac{\exp(f_i)}{\exp(f_i^*)} \right) = \sum_{k}^{K} (G_{ki} - G_{ki}^*) \theta_k.
\]

(Snijders, Van de Bunt and Steglich, 2010)
SAOM: rate function

The rate function is similarly a function of network summary statistics from the node perspective:

\[ r_i = \sum_{k}^{K} G_{ki} \gamma_k, \]

but note that these might be, and typically are, different characteristics.

Since the rate is positive, this is modelled using an exponential distribution, while the objective function is the input of a multinomial distribution over all possible changes.

An example is the coevolution of a friendship network and delinquency. Note the presence of:

- Network structural effects.
- Node attribute impacts on ties.
- Friends’ node attribute impacts on ties.
- Node attribute impacts on rate of change.
- Rate changes for ties and nodes, separately.

(Snijders, Van de Bunt and Steglich, 2010, 47)
Example summary statistics

“Average alter effect” (avAlt) measure:

\[ s_i^{avAlt}(W, y) = y_i \cdot \frac{\sum_j w_{ij} y_j}{\sum_j w_{ij}}. \]

“Average similarity effect” (avSim) measure:

\[ s_i^{avSim}(W, y) = \frac{\sum_j w_{ij} (\delta_{ij}^y - \hat{\delta}^y)}{\sum_j w_{ij}}, \]

where \( s_i^{avSim} = 0 \) if \( \sum_j w_{ij} = 0 \), with

\[ \delta_{ij}^y = \delta(y_i, y_j) = 1 - |y_i - y_j| \]

a similarity measure of \( y_i \) and \( y_j \) and \( \hat{\delta}^y \) the mean similarity score across the network.

(Ripley et al., 2018)
Estimate of $\rho$ (left) and power of associated $t$-tests (right), when estimating spatial clustering parameter using static SAOM specification, using:

$$s_i^{avSim}(W, y) = \frac{\sum_j w_{ij}(\delta_{ij}^y - \hat{\delta}_y)}{\sum_j w_{ij}},$$

(Elkink and Grund, 2018)
Coevolution

Much of current developments are around coevolution, the idea that both spatial and network effects operate simultaneously and affect each other.

For example: countries that trade with democracies are more likely to become democratic, and democracies are more likely to trade with each other.

(Kachi, 2012; Mohrenberg, 2013)
Further reading

Good texts as a starting point would be:

Spatial: Anselin (2002); Ward and Gleditsch (2008); Darmofal (2015). LeSage and Pace (2009) is a bit more advanced and the classic is Anselin (1988).

ERGM: Cranmer and Desmarais (2011); Lusher, Koskinen and Robins (2012).

SAOM: Snijders, Van de Bunt and Steglish (2010).


URL: [http://dx.doi.org/10.1068/a231025](http://dx.doi.org/10.1068/a231025)


