Exponential Family Random Graph Model (ERGM)

Definition

$$\Pr(\mathbf{Y} = \mathbf{y}; \boldsymbol{\theta}) = \frac{e^{\boldsymbol{\theta} \cdot g(\mathbf{y})}}{k(\boldsymbol{\theta})}, \ \mathbf{y} \in \mathcal{Y},$$

$$k(\boldsymbol{\theta}) = \sum_{\mathbf{y}' \in \mathcal{Y}} e^{\boldsymbol{\theta} \cdot g(\mathbf{y}')}$$

- $\boldsymbol{\theta}$ a vector of model parameters
- $g(\cdot)$ a vector of sufficient statistics (also incorporating exogenous information)
- $k(\cdot)$ the normalizing constant, often intractable

Conditional log-odds of a tie

The local view of ERGM

$$\text{logit}\left(\Pr(\mathbf{Y}_{i,j} = 1 | \mathbf{Y}_{-(i,j)} = \mathbf{y}_{-(i,j)}; \boldsymbol{\theta})\right)$$

Conditional log-odds of a tie

The local view of ERGM

$$\text{logit}\left(\Pr(\mathbf{Y}_{i,j} = 1 | \mathbf{Y}_{-(i,j)} = \mathbf{y}_{-(i,j)}; \boldsymbol{\theta})\right)$$
$$= \log \frac{\Pr(\mathbf{Y} = \mathbf{y}_{+(i,j)}; \boldsymbol{\theta})}{\Pr(\mathbf{Y} = \mathbf{y}_{-(i,j)}; \boldsymbol{\theta})}$$

Conditional log-odds of a tie

The local view of ERGM

$$\operatorname{logit}\left(\Pr(\mathbf{Y}_{i,j} = 1 | \mathbf{Y}_{-(i,j)} = \mathbf{y}_{-(i,j)}; \boldsymbol{\theta})\right)$$
$$= \log \frac{\Pr(\mathbf{Y} = \mathbf{y}_{+(i,j)}; \boldsymbol{\theta})}{\Pr(\mathbf{Y} = \mathbf{y}_{-(i,j)}; \boldsymbol{\theta})}$$

Conditional log-odds of a tie

The local view of ERGM

$$\text{logit}\left(\Pr(\mathbf{Y}_{i,j} = 1 | \mathbf{Y}_{-(i,j)} = \mathbf{y}_{-(i,j)}; \boldsymbol{\theta})\right)$$
$$= \log\left(\frac{e^{\boldsymbol{\theta}\cdot g(\mathbf{y}_{+(i,j)})}}{\cancel{k(\boldsymbol{\theta})}} \div \frac{e^{\boldsymbol{\theta}\cdot g(\mathbf{y}_{-(i,j)})}}{\cancel{k(\boldsymbol{\theta})}}\right)$$

Conditional log-odds of a tie

The local view of ERGM

$$\text{logit}\left(\Pr(\mathbf{Y}_{i,j} = 1 | \mathbf{Y}_{-(i,j)} = \mathbf{y}_{-(i,j)}; \boldsymbol{\theta})\right)$$
$$= \boldsymbol{\theta} \cdot \delta(\mathbf{y}_{i,j})$$

Conditional log-odds of a tie

The local view of ERGM

$$\text{logit}\left(\Pr(\mathbf{Y}_{i,j} = 1 | \mathbf{Y}_{-(i,j)} = \mathbf{y}_{-(i,j)}; \boldsymbol{\theta})\right)$$
$$= \boldsymbol{\theta} \cdot \delta(\mathbf{y}_{i,j})$$

Conditional log-odds of a tie

The local view of ERGM

$$\text{logit}\left(\Pr(\mathbf{Y}_{i,j} = 1 | \mathbf{Y}_{-(i,j)} = \mathbf{y}_{-(i,j)}; \boldsymbol{\theta})\right)$$
$$= \boldsymbol{\theta} \cdot \delta(\mathbf{y}_{i,j})$$

$$\delta(\mathbf{y}_{i,j}) \stackrel{\text{def}}{=} g(\mathbf{y}_{+(i,j)}) - g(\mathbf{y}_{-(i,j)})$$

Conditional log-odds of a tie

The local view of ERGM

$$\text{logit}\left(\Pr(\mathbf{Y}_{i,j} = 1 | \mathbf{Y}_{-(i,j)} = \mathbf{y}_{-(i,j)}; \boldsymbol{\theta})\right)$$
$$= \boldsymbol{\theta} \cdot \delta(\mathbf{y}_{i,j})$$

$$\delta(\mathbf{y}_{i,j}) \stackrel{\text{def}}{=} g(\mathbf{y}_{+(i,j)}) - g(\mathbf{y}_{-(i,j)})$$

Conditional log-odds of a tie

The local view of ERGM

$$\begin{aligned}
&\text{logit}\left(\Pr(\mathbf{Y}_{i,j} = 1 | \mathbf{Y}_{-(i,j)} = \mathbf{y}_{-(i,j)}; \boldsymbol{\theta})\right) \\
&\quad = \boldsymbol{\theta} \cdot \delta(\mathbf{y}_{i,j})
\end{aligned}$$

$$\delta(\mathbf{y}_{i,j}) \stackrel{\text{def}}{=} g(\mathbf{y}_{+(i,j)}) - g(\mathbf{y}_{-(i,j)})$$

Prevalence, incidence, and duration

$$\text{Prevalence} = \text{Incidence} \times \text{Duration}$$

Prevalence, incidence, and duration

$$\text{Prevalence} \quad = \quad \text{Incidence} \quad \times \quad \text{Duration}$$

Prevalence, incidence, and duration

$$\text{Prevalence} = \text{Incidence} \times \text{Duration}$$

Why separate formation from dissolution?

Intuition The social forces that facilitate formation of ties are often different from those that facilitate their dissolution.

Interpretation Because of this, we would want model parameters to be interpreted in terms of ties formed and ties dissolved.

Inference We need to fit these models to a single network or ego-centric data, and incorporate dynamics separately.

Confounding We want to control confounding between incidence and duration.

Why separate formation from dissolution?

Intuition The social forces that facilitate formation of ties are often different from those that facilitate their dissolution.

Interpretation Because of this, we would want model parameters to be interpreted in terms of ties formed and ties dissolved.

Inference We need to fit these models to a single network or ego-centric data, and incorporate dynamics separately.

Confounding We want to control confounding between incidence and duration.

Why separate formation from dissolution?

Intuition The social forces that facilitate formation of ties are often different from those that facilitate their dissolution.

Interpretation Because of this, we would want model parameters to be interpreted in terms of ties formed and ties dissolved.

Inference We need to fit these models to a single network or ego-centric data, and incorporate dynamics separately.

Confounding We want to control confounding between incidence and duration.

Why separate formation from dissolution?

Intuition The social forces that facilitate formation of ties are often different from those that facilitate their dissolution.

Interpretation Because of this, we would want model parameters to be interpreted in terms of ties formed and ties dissolved.

Inference We need to fit these models to a single network or ego-centric data, and incorporate dynamics separately.

Confounding We want to control confounding between incidence and duration.

Confounding problem

$$\text{Prevalence} \quad = \quad \text{Incidence} \quad \times \quad \text{Duration}$$

Confounding problem

$$\underline{\text{Prevalence}} = \text{Incidence} \times \text{Duration}$$

Confounding problem

$$\begin{array}{ccccc} \text{low} & \Rightarrow & \text{rare?} & \textit{or} & \text{short?} \\ \text{Prevalence} & = & \text{Incidence} & \times & \text{Duration} \end{array}$$

Idea

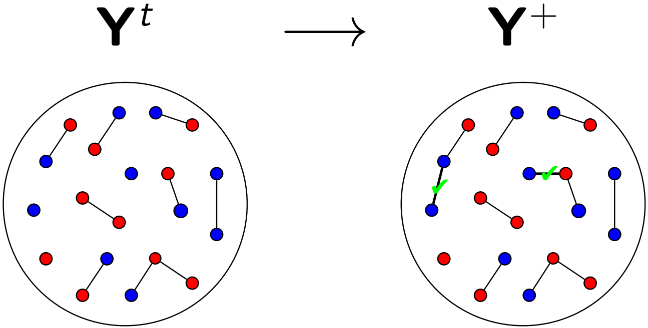
Represent evolution from $\mathbf{Y}^t$ to $\mathbf{Y}^{t+1}$ as a product of two phases: one in which ties are formed and another in which they are dissolved, each phase a draw from an ERGM.

Idea

Represent evolution from $\mathbf{Y}^t$ to $\mathbf{Y}^{t+1}$ as a product of two phases: one in which ties are formed and another in which they are dissolved, each phase a draw from an ERGM.

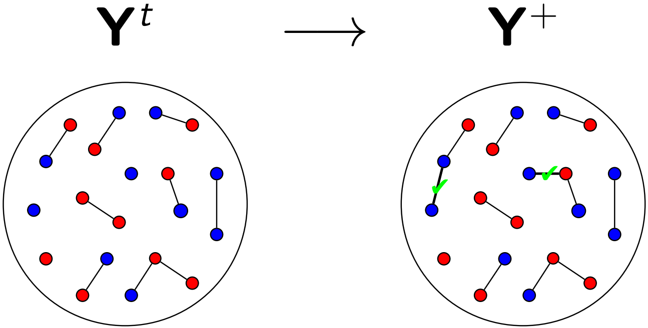
Formation Phase

$$\Pr(\mathbf{Y}^+ = \mathbf{y}^+ | \mathbf{Y}^t = \mathbf{y}^t; \boldsymbol{\theta}^+) = \frac{e^{\boldsymbol{\theta}^+ \cdot g^+(\mathbf{y}^+, \mathbf{y}^t)} 1_{\mathbf{y}^+ \supseteq \mathbf{y}^t}}{k^+(\boldsymbol{\theta}^+, \mathbf{y}^t)}, \ \mathbf{y}^+ \in \mathcal{Y}$$



Formation Phase

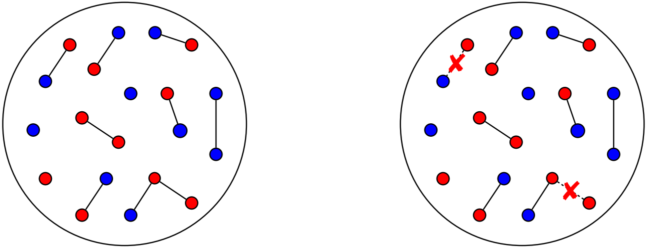
$$\Pr(\mathbf{Y}^+ = \mathbf{y}^+ | \mathbf{Y}^t = \mathbf{y}^t; \boldsymbol{\theta}^+) = \frac{e^{\boldsymbol{\theta}^+ \cdot g^+(\mathbf{y}^+, \mathbf{y}^t)} \mathbb{1}_{\mathbf{y}^+ \supseteq \mathbf{y}^t}}{k^+(\boldsymbol{\theta}^+, \mathbf{y}^t)}, \ \mathbf{y}^+ \in \mathcal{Y}$$



Dissolution/Preservation Phase

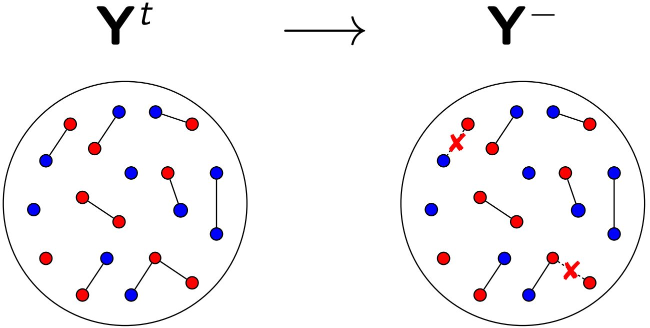
$$\Pr(\mathbf{Y}^- = \mathbf{y}^- | \mathbf{Y}^t = \mathbf{y}^t; \boldsymbol{\theta}^-) = \frac{e^{\boldsymbol{\theta}^- \cdot g^-(\mathbf{y}^-, \mathbf{y}^t)} 1_{\mathbf{y}^- \subseteq \mathbf{y}^t}}{k^-(\boldsymbol{\theta}^-, \mathbf{y}^t)}, \ \mathbf{y}^- \in \mathcal{Y}$$

$$\mathbf{Y}^t \longrightarrow \mathbf{Y}^-$$

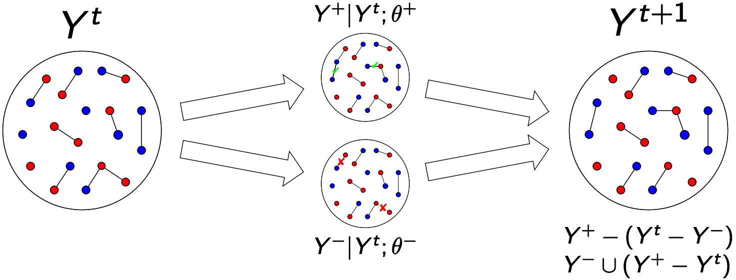


Dissolution/Preservation Phase

$$\Pr(\mathbf{Y}^- = \mathbf{y}^- | \mathbf{Y}^t = \mathbf{y}^t; \boldsymbol{\theta}^-) = \frac{e^{\boldsymbol{\theta}^- \cdot g^-(\mathbf{y}^-, \mathbf{y}^t)} \color{red}{1_{\mathbf{y}^- \subseteq \mathbf{y}^t}}}{k^-(\boldsymbol{\theta}^-, \mathbf{y}^t)}, \ \mathbf{y}^- \in \mathcal{Y}$$



Putting the Two Together



Effect of terms

Tie count

Let $g_k(\mathbf{y}) = |\mathbf{y}|$. Then, other things being equal...

	$\theta \nearrow$	$\theta \searrow$
formation phase	more new ties created each time step	fewer new ties created each time step
dissolution/ preservation phase	more existing ties preserved (fewer dissolved); longer average duration	fewer existing ties preserved (more dissolved); shorter average duration

Effect of terms

Tie count

Let $g_k(\mathbf{y}) = |\mathbf{y}|$. Then, other things being equal...

	$\boldsymbol{\theta} \nearrow$	$\boldsymbol{\theta} \searrow$
formation phase	more new ties created each time step	fewer new ties created each time step
dissolution/ preservation phase	more existing ties preserved (fewer dissolved); longer average duration	fewer existing ties preserved (more dissolved); shorter average duration

Effect of terms

Tie count

Let $g_k(\mathbf{y}) = |\mathbf{y}|$. Then, other things being equal...

	$\boldsymbol{\theta} \nearrow$	$\boldsymbol{\theta} \searrow$
formation phase	more new ties created each time step	fewer new ties created each time step
dissolution/ preservation phase	more existing ties preserved (fewer dissolved); longer average duration	fewer existing ties preserved (more dissolved); shorter average duration

Effect of terms

Tie count

Let $g_k(\mathbf{y}) = |\mathbf{y}|$. Then, other things being equal...

	$\boldsymbol{\theta} \nearrow$	$\boldsymbol{\theta} \searrow$
formation phase	more new ties created each time step	fewer new ties created each time step
dissolution/ preservation phase	more existing ties preserved (fewer dissolved); longer average duration	fewer existing ties preserved (more dissolved); shorter average duration

Effect of terms

Tie count

Let $g_k(\mathbf{y}) = |\mathbf{y}|$. Then, other things being equal...

	$\boldsymbol{\theta} \nearrow$	$\boldsymbol{\theta} \searrow$
formation phase	more new ties created each time step	fewer new ties created each time step
dissolution/ preservation phase	more existing ties preserved (fewer dissolved); longer average duration	fewer existing ties preserved (more dissolved); shorter average duration

Effect of terms

Tie count

Let $g_k(\mathbf{y}) = |\mathbf{y}|$. Then, other things being equal...

	$\boldsymbol{\theta} \nearrow$	$\boldsymbol{\theta} \searrow$
formation phase	more new ties created each time step	fewer new ties created each time step
dissolution/ preservation phase	more existing ties preserved (fewer dissolved); longer average duration	fewer existing ties preserved (more dissolved); shorter average duration

Effect of terms

Tie count

Let $g_k(\mathbf{y}) = |\mathbf{y}|$. Then, other things being equal...

	$\boldsymbol{\theta} \nearrow$	$\boldsymbol{\theta} \searrow$
formation phase	more new ties created each time step	fewer new ties created each time step
dissolution/ preservation phase	more existing ties preserved (fewer dissolved); longer average duration	fewer existing ties preserved (more dissolved); shorter average duration

Since the network evolves slowly...

Effect of terms

Degree distribution

Let $g_k(\mathbf{y}) = \sum_i 1_{|\mathbf{y}_i| \geq 2}$. Then, other things being equal...

	$\theta \nearrow$	$\theta \searrow$
formation phase	more ties added to actors with exactly 1 tie	fewer ties added to actors with 1 tie
dissolution/ preservation phase	actor's second tie more likely to be preserved	actor's second tie more likely to be dissolved

Effect of terms

Degree distribution

Let $g_k(\mathbf{y}) = \sum_i 1_{|\mathbf{y}_i| \geq 2}$. Then, other things being equal...

	$\boldsymbol{\theta} \nearrow$	$\boldsymbol{\theta} \searrow$
formation phase	more ties added to actors with exactly 1 tie	fewer ties added to actors with 1 tie
dissolution/ preservation phase	actor's second tie more likely to be preserved	actor's second tie more likely to be dissolved

Effect of terms

Degree distribution

Let $g_k(\mathbf{y}) = \sum_i 1_{|\mathbf{y}_i| \geq 2}$. Then, other things being equal...

	$\boldsymbol{\theta} \nearrow$	$\boldsymbol{\theta} \searrow$
formation phase	more ties added to actors with exactly 1 tie	fewer ties added to actors with 1 tie
dissolution/ preservation phase	actor's second tie more likely to be preserved	actor's second tie more likely to be dissolved

Effect of terms

Degree distribution

Let $g_k(\mathbf{y}) = \sum_i 1_{|\mathbf{y}_i| \geq 2}$. Then, other things being equal...

	$\boldsymbol{\theta} \nearrow$	$\boldsymbol{\theta} \searrow$
formation phase	more ties added to actors with exactly 1 tie	fewer ties added to actors with 1 tie
dissolution/ preservation phase	actor's second tie more likely to be preserved	actor's second tie more likely to be dissolved

Effect of terms

Degree distribution

Let $g_k(\mathbf{y}) = \sum_i 1_{|\mathbf{y}_i| \geq 2}$. Then, other things being equal...

	$\boldsymbol{\theta} \nearrow$	$\boldsymbol{\theta} \searrow$
formation phase	more ties added to actors with exactly 1 tie	fewer ties added to actors with 1 tie
dissolution/ preservation phase	actor's second tie more likely to be preserved	actor's second tie more likely to be dissolved

Effect of terms

Degree distribution

Let $g_k(\mathbf{y}) = \sum_i 1_{|\mathbf{y}_i| \geq 2}$. Then, other things being equal...

	$\boldsymbol{\theta} \nearrow$	$\boldsymbol{\theta} \searrow$
formation phase	more ties added to actors with exactly 1 tie	fewer ties added to actors with 1 tie
dissolution/ preservation phase	actor's second tie more likely to be preserved	actor's second tie more likely to be dissolved

Effect of terms

Degree distribution

Let $g_k(\mathbf{y}) = \sum_i 1_{|\mathbf{y}_i| \geq 2}$. Then, other things being equal...

	$\boldsymbol{\theta} \nearrow$	$\boldsymbol{\theta} \searrow$
formation phase	more ties added to actors with exactly 1 tie	fewer ties added to actors with 1 tie
dissolution/ preservation phase	actor's second tie more likely to be preserved	actor's second tie more likely to be dissolved

To encourage "monogamy"...

Equilibrium

- In the long run, network evolution converges to an equilibrium distribution $\Pr_g(\mathbf{Y} = \mathbf{y}; \boldsymbol{\theta}^+, \boldsymbol{\theta}^-)$.
- We can treat an observed network as a random draw from this distribution.

Equilibrium

- In the long run, network evolution converges to an equilibrium distribution $\Pr_g(\mathbf{Y} = \mathbf{y}; \boldsymbol{\theta}^+, \boldsymbol{\theta}^-)$.
- We can treat an observed network as a random draw from this distribution.

Back to Prevalence

$$\text{Prevalence} \quad = \quad \text{Incidence} \quad \times \quad \text{Duration}$$

Back to Prevalence

$$\begin{array}{ccccc} \text{Prevalence} & = & \text{Incidence} & \times & \text{Duration} \\ & & \| & & \\ & & \text{Formation} & & \end{array}$$

Back to Prevalence

$$\text{Prevalence} = \text{Incidence} \times \underset{\text{Dissolution}}{\underset{||}{\text{Duration}}}$$

Back to Prevalence

$$\begin{array}{ccccc} \text{Prevalence} & = & \text{Incidence} & \times & \text{Duration} \\ \| & & & & \\ \text{Equilibrium} & & & & \end{array}$$

Back to Prevalence

$$\begin{array}{ccccc} \text{Prevalence} & = & \text{Incidence} & \times & \text{Duration} \\ \| & & \| & & \| \\ \text{Equilibrium} & & \text{Formation} & & \text{Dissolution} \end{array}$$