

# **Modeling Networks with Missing Data in statnet**

## **ERGM Refresher**

**Carter T. Butts**

Department of Sociology and  
Institute for Mathematical Behavioral Sciences  
University of California, Irvine

2011 Political Networks Conference

June 16, 2011

Ann Arbor, MI

<http://polnet2011.statnet.org/>

# Exponential Families for Random Graphs (w/Covariates)

- For random graph  $G$  w/countable support  $\mathcal{G}$  and covariate set  $X$ , pmf can be written in ERG form:

$$\Pr(G = g | \mathbf{t}, \theta, \mathcal{G}, X) = \frac{\exp(\theta^T \mathbf{t}(g, X))}{\sum_{g' \in \mathcal{G}} \exp(\theta^T \mathbf{t}(g', X))} I_{\mathcal{G}}(g)$$

- $\theta^T \mathbf{t}$ : linear predictor
  - $\mathbf{t}: \mathcal{G} \rightarrow \mathbb{R}^m$ : vector of sufficient statistics
  - $\theta \in \mathbb{R}^m$ : vector of parameters
  - $\sum_{g' \in \mathcal{G}} \exp(\theta^T \mathbf{t}(g', X))$ : normalizing factor
- **Intuition: ERG placed more/less weight on structures with certain features, as determined by  $\mathbf{t}, \theta$** 
  - Framework is complete for pmfs on  $\mathcal{G}$ , few constraints on  $\mathbf{t}$

# Exponential Families for Random Graphs (w/Covariates)

- For random graph  $G$  w/countable support  $\mathcal{G}$  and covariate set  $X$ , pmf can be written in ERG form:

$$\Pr(G = g | \mathbf{t}, \theta, \mathcal{G}, X) = \frac{\exp(\theta^T \mathbf{t}(g, X))}{\sum_{g' \in \mathcal{G}} \exp(\theta^T \mathbf{t}(g', X))} I_{\mathcal{G}}(g)$$

**Source of great difficulty!**



- $\theta^T \mathbf{t}$ : linear predictor
  - $\mathbf{t}: \mathcal{G} \rightarrow \mathbb{R}^m$ : vector of sufficient statistics
  - $\theta \in \mathbb{R}^m$ : vector of parameters
  - $\sum_{g' \in \mathcal{G}} \exp(\theta^T \mathbf{t}(g', X))$ : normalizing factor
- **Intuition: ERG placed more/less weight on structures with certain features, as determined by  $\mathbf{t}, \theta$** 
  - Framework is complete for pmfs on  $\mathcal{G}$ , few constraints on  $\mathbf{t}$

# Equivalent Expression: Modeling the Adjacency Matrix

- For adjacency matrix  $Y$  w/countable support  $\mathcal{Y}$  and covariate set  $X$ , any pmf can be written in ERG form:

$$\Pr(\mathbf{Y}=\mathbf{y}|\mathbf{t}, \theta, \mathcal{Y}, X) = \frac{\exp(\theta^T \mathbf{t}(\mathbf{y}, X))}{\sum_{\mathbf{y}' \in \mathcal{Y}} \exp(\theta^T \mathbf{t}(\mathbf{y}', X))} I_{\mathcal{Y}}(\mathbf{y})$$

- $\theta^T \mathbf{t}$ : linear predictor
  - $\mathbf{t}: \mathcal{Y} \rightarrow \mathbb{R}^m$ : vector of sufficient statistics
  - $\theta \in \mathbb{R}^m$ : vector of parameters
  - $\sum_{\mathbf{y}' \in \mathcal{Y}} \exp(\theta^T \mathbf{t}(\mathbf{y}', X))$ : normalizing factor
- **Additional notation:**
  - $\mathbf{y}_{ij}^+, \mathbf{y}_{ij}^-$ :  $y$  w/ $ij$ th cell set to 1 or 0 (respectively)
  - $\mathbf{y}_{ij}^c$ : all elements of  $y$  other than the  $ij$ th

# ERGs and Conditional Odds of an Edge

- Can easily specify the conditional odds of an edge:

$$\frac{\Pr(\mathbf{Y} = \mathbf{y}_{ij}^+ | \mathbf{t}, \theta, \mathcal{Y}, X)}{\Pr(\mathbf{Y} = \mathbf{y}_{ij}^- | \mathbf{t}, \theta, \mathcal{Y}, X)} = \frac{\exp(\theta^T \mathbf{t}(\mathbf{y}_{ij}^+, X)) \sum_{\mathbf{y}' \in \mathcal{Y}} \exp(\theta^T \mathbf{t}(\mathbf{y}', X))}{\sum_{\mathbf{y}' \in \mathcal{Y}} \exp(\theta^T \mathbf{t}(\mathbf{y}', X)) \exp(\theta^T \mathbf{t}(\mathbf{y}_{ij}^-, X))}$$

# ERGs and Conditional Odds of an Edge

- Can easily specify the conditional odds of an edge:

$$\begin{aligned} \frac{\Pr(\mathbf{Y} = \mathbf{y}_{ij}^+ | \mathbf{t}, \theta, \mathcal{Y}, X)}{\Pr(\mathbf{Y} = \mathbf{y}_{ij}^- | \mathbf{t}, \theta, \mathcal{Y}, X)} &= \frac{\exp(\theta^T \mathbf{t}(\mathbf{y}_{ij}^+, X)) \sum_{\mathbf{y}' \in \mathcal{Y}} \exp(\theta^T \mathbf{t}(\mathbf{y}', X))}{\sum_{\mathbf{y}' \in \mathcal{Y}} \exp(\theta^T \mathbf{t}(\mathbf{y}', X)) \exp(\theta^T \mathbf{t}(\mathbf{y}_{ij}^-, X))} \\ &= \frac{\exp(\theta^T \mathbf{t}(\mathbf{y}_{ij}^+, X))}{\exp(\theta^T \mathbf{t}(\mathbf{y}_{ij}^-, X))} = \exp(\theta^T (\mathbf{t}(\mathbf{y}_{ij}^+, X) - \mathbf{t}(\mathbf{y}_{ij}^-, X))) \end{aligned}$$

# ERGs and Conditional Odds of an Edge

- Can easily specify the conditional odds of an edge:

$$\frac{\Pr(\mathbf{Y} = \mathbf{y}_{ij}^+ | \mathbf{t}, \theta, \mathcal{Y}, X)}{\Pr(\mathbf{Y} = \mathbf{y}_{ij}^- | \mathbf{t}, \theta, \mathcal{Y}, X)} = \frac{\exp(\theta^T \mathbf{t}(\mathbf{y}_{ij}^+, X)) \sum_{\mathbf{y}' \in \mathcal{Y}} \exp(\theta^T \mathbf{t}(\mathbf{y}', X))}{\sum_{\mathbf{y}' \in \mathcal{Y}} \exp(\theta^T \mathbf{t}(\mathbf{y}', X)) \exp(\theta^T \mathbf{t}(\mathbf{y}_{ij}^-, X))} = \frac{\exp(\theta^T \mathbf{t}(\mathbf{y}_{ij}^+, X))}{\exp(\theta^T \mathbf{t}(\mathbf{y}_{ij}^-, X))} = \exp(\theta^T (\mathbf{t}(\mathbf{y}_{ij}^+, X) - \mathbf{t}(\mathbf{y}_{ij}^-, X)))$$

- Log-odds depend on the changescore,  $\Delta_{ij} = \mathbf{t}(\mathbf{y}_{ij}^+, X) - \mathbf{t}(\mathbf{y}_{ij}^-, X)$
- Useful implication: each unit change in  $\mathbf{t}_k$  for  $(i,j)$  edge present (versus absent) increases the conditional log-odds of  $(i,j)$  by  $\theta_k$
- **Important:** this is only conditionally true! The marginal log-odds of an  $(i,j)$  edge can depend on a complex way on other aspects of the graph

# Conditional Edge Probability

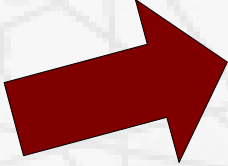
- From the above, we can also determine conditional edge probability using fact that  $\text{prob} = \text{odds} / (1 + \text{odds})$ :

$$\Pr(Y_{ij} = 1 | Y_{ij}^c = y_{ij}^c, t, \theta, \mathcal{Y}, X) = \Pr(Y = y_{ij}^+ | Y_{ij}^c = y_{ij}^c, t, \theta, \mathcal{Y}, X)$$



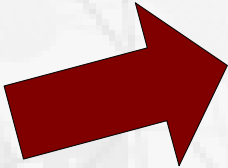
# Conditional Edge Probability

- From the above, we can also determine conditional edge probability using fact that  $\text{prob} = \text{odds} / (1 + \text{odds})$ :

$$\begin{aligned} \Pr(\mathbf{Y}_{ij} = 1 \mid \mathbf{Y}_{ij}^c = \mathbf{y}_{ij}^c, \mathbf{t}, \theta, \mathcal{Y}, X) &= \Pr(\mathbf{Y} = \mathbf{y}_{ij}^+ \mid \mathbf{Y}_{ij}^c = \mathbf{y}_{ij}^c, \mathbf{t}, \theta, \mathcal{Y}, X) \\ &= \frac{\Pr(\mathbf{Y} = \mathbf{y}_{ij}^+ \mid \mathbf{t}, \theta, \mathcal{Y}, X)}{\Pr(\mathbf{Y} = \mathbf{y}_{ij}^- \mid \mathbf{t}, \theta, \mathcal{Y}, X)} \left[ 1 + \frac{\Pr(\mathbf{Y} = \mathbf{y}_{ij}^+ \mid \mathbf{t}, \theta, \mathcal{Y}, X)}{\Pr(\mathbf{Y} = \mathbf{y}_{ij}^- \mid \mathbf{t}, \theta, \mathcal{Y}, X)} \right]^{-1} \end{aligned}$$


# Conditional Edge Probability

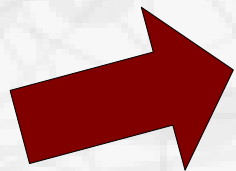
- From the above, we can also determine conditional edge probability using fact that  $\text{prob} = \text{odds} / (1 + \text{odds})$ :

$$\begin{aligned} \Pr(\mathbf{Y}_{ij} = 1 \mid \mathbf{Y}_{ij}^c = \mathbf{y}_{ij}^c, \mathbf{t}, \theta, \mathcal{Y}, X) &= \Pr(\mathbf{Y} = \mathbf{y}_{ij}^+ \mid \mathbf{Y}_{ij}^c = \mathbf{y}_{ij}^c, \mathbf{t}, \theta, \mathcal{Y}, X) \\ &= \frac{\Pr(\mathbf{Y} = \mathbf{y}_{ij}^+ \mid \mathbf{t}, \theta, \mathcal{Y}, X)}{\Pr(\mathbf{Y} = \mathbf{y}_{ij}^- \mid \mathbf{t}, \theta, \mathcal{Y}, X)} \left[ 1 + \frac{\Pr(\mathbf{Y} = \mathbf{y}_{ij}^+ \mid \mathbf{t}, \theta, \mathcal{Y}, X)}{\Pr(\mathbf{Y} = \mathbf{y}_{ij}^- \mid \mathbf{t}, \theta, \mathcal{Y}, X)} \right]^{-1} \\ &= \frac{\exp\left(\theta^T \left(\mathbf{t}(\mathbf{y}_{ij}^+, X) - \mathbf{t}(\mathbf{y}_{ij}^-, X)\right)\right)}{1 + \exp\left(\theta^T \left(\mathbf{t}(\mathbf{y}_{ij}^+, X) - \mathbf{t}(\mathbf{y}_{ij}^-, X)\right)\right)} \end{aligned}$$


# Conditional Edge Probability

- From the above, we can also determine conditional edge probability using fact that  $\text{prob} = \text{odds} / (1 + \text{odds})$ :

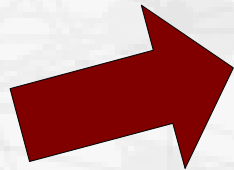
$$\begin{aligned} \Pr(\mathbf{Y}_{ij} = 1 \mid \mathbf{Y}_{ij}^c = \mathbf{y}_{ij}^c, \mathbf{t}, \theta, \mathcal{Y}, X) &= \Pr(\mathbf{Y} = \mathbf{y}_{ij}^+ \mid \mathbf{Y}_{ij}^c = \mathbf{y}_{ij}^c, \mathbf{t}, \theta, \mathcal{Y}, X) \\ &= \frac{\Pr(\mathbf{Y} = \mathbf{y}_{ij}^+ \mid \mathbf{t}, \theta, \mathcal{Y}, X)}{\Pr(\mathbf{Y} = \mathbf{y}_{ij}^- \mid \mathbf{t}, \theta, \mathcal{Y}, X)} \left[ 1 + \frac{\Pr(\mathbf{Y} = \mathbf{y}_{ij}^+ \mid \mathbf{t}, \theta, \mathcal{Y}, X)}{\Pr(\mathbf{Y} = \mathbf{y}_{ij}^- \mid \mathbf{t}, \theta, \mathcal{Y}, X)} \right]^{-1} \\ &= \frac{\exp\left(\theta^T (\mathbf{t}(\mathbf{y}_{ij}^+, X) - \mathbf{t}(\mathbf{y}_{ij}^-, X))\right)}{1 + \exp\left(\theta^T (\mathbf{t}(\mathbf{y}_{ij}^+, X) - \mathbf{t}(\mathbf{y}_{ij}^-, X))\right)} \\ &= \left[ 1 + \exp\left(\theta^T (\mathbf{t}(\mathbf{y}_{ij}^-, X) - \mathbf{t}(\mathbf{y}_{ij}^+, X))\right) \right]^{-1} \end{aligned}$$



# Conditional Edge Probability

- From the above, we can also determine conditional edge probability using fact that  $\text{prob} = \text{odds} / (1 + \text{odds})$ :

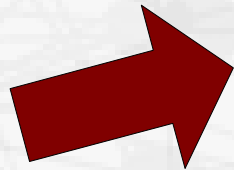
$$\begin{aligned} \Pr(\mathbf{Y}_{ij} = 1 \mid \mathbf{Y}_{ij}^c = \mathbf{y}_{ij}^c, \mathbf{t}, \theta, \mathcal{Y}, X) &= \Pr(\mathbf{Y} = \mathbf{y}_{ij}^+ \mid \mathbf{Y}_{ij}^c = \mathbf{y}_{ij}^c, \mathbf{t}, \theta, \mathcal{Y}, X) \\ &= \frac{\Pr(\mathbf{Y} = \mathbf{y}_{ij}^+ \mid \mathbf{t}, \theta, \mathcal{Y}, X)}{\Pr(\mathbf{Y} = \mathbf{y}_{ij}^- \mid \mathbf{t}, \theta, \mathcal{Y}, X)} \left[ 1 + \frac{\Pr(\mathbf{Y} = \mathbf{y}_{ij}^+ \mid \mathbf{t}, \theta, \mathcal{Y}, X)}{\Pr(\mathbf{Y} = \mathbf{y}_{ij}^- \mid \mathbf{t}, \theta, \mathcal{Y}, X)} \right]^{-1} \\ &= \frac{\exp(\theta^T (\mathbf{t}(\mathbf{y}_{ij}^+, X) - \mathbf{t}(\mathbf{y}_{ij}^-, X)))}{1 + \exp(\theta^T (\mathbf{t}(\mathbf{y}_{ij}^+, X) - \mathbf{t}(\mathbf{y}_{ij}^-, X)))} \\ &= \left[ 1 + \exp(\theta^T (\mathbf{t}(\mathbf{y}_{ij}^-, X) - \mathbf{t}(\mathbf{y}_{ij}^+, X))) \right]^{-1} \\ &= \text{logit}^{-1}(\theta^T (\mathbf{t}(\mathbf{y}_{ij}^+, X) - \mathbf{t}(\mathbf{y}_{ij}^-, X))) \end{aligned}$$



# Conditional Edge Probability

- From the above, we can also determine conditional edge probability using fact that  $\text{prob} = \text{odds} / (1 + \text{odds})$ :

$$\begin{aligned} \Pr(\mathbf{Y}_{ij} = 1 \mid \mathbf{Y}_{ij}^c = \mathbf{y}_{ij}^c, \mathbf{t}, \theta, \mathcal{Y}, X) &= \Pr(\mathbf{Y} = \mathbf{y}_{ij}^+ \mid \mathbf{Y}_{ij}^c = \mathbf{y}_{ij}^c, \mathbf{t}, \theta, \mathcal{Y}, X) \\ &= \frac{\Pr(\mathbf{Y} = \mathbf{y}_{ij}^+ \mid \mathbf{t}, \theta, \mathcal{Y}, X)}{\Pr(\mathbf{Y} = \mathbf{y}_{ij}^- \mid \mathbf{t}, \theta, \mathcal{Y}, X)} \left[ 1 + \frac{\Pr(\mathbf{Y} = \mathbf{y}_{ij}^+ \mid \mathbf{t}, \theta, \mathcal{Y}, X)}{\Pr(\mathbf{Y} = \mathbf{y}_{ij}^- \mid \mathbf{t}, \theta, \mathcal{Y}, X)} \right]^{-1} \\ &= \frac{\exp(\theta^T (\mathbf{t}(\mathbf{y}_{ij}^+, X) - \mathbf{t}(\mathbf{y}_{ij}^-, X)))}{1 + \exp(\theta^T (\mathbf{t}(\mathbf{y}_{ij}^+, X) - \mathbf{t}(\mathbf{y}_{ij}^-, X)))} \\ &= \left[ 1 + \exp(\theta^T (\mathbf{t}(\mathbf{y}_{ij}^-, X) - \mathbf{t}(\mathbf{y}_{ij}^+, X))) \right]^{-1} \\ &= \text{logit}^{-1}(\theta^T (\mathbf{t}(\mathbf{y}_{ij}^+, X) - \mathbf{t}(\mathbf{y}_{ij}^-, X))) \end{aligned}$$



# Conditional Edge Probability, Cont.

- **So, the conditional probability of an  $(i,j)$  edge is simply the inverse logit of  $\theta^T \Delta_{ij}$** 
  - Obvious idea: to find  $\theta$ , why not set this up as a logistic network regression problem (regressing  $y$  on  $\Delta$ )?
    - This is an “autologistic regression,” and the resulting estimator is known as a *pseudolikelihood* estimator (Besag, 1975)
  - Problem: the probability here is only conditional – can use for any one  $ij$ , but joint likelihood of  $y$  is not generally the product of  $\Pr(\mathbf{Y}_{ij} = \mathbf{y}_{ij} | \mathbf{Y}_{ij}^c = \mathbf{y}_{ij}^c)$ 
    - Another view:  $y$  appears on both sides – can't regress w/out accounting for the “feedback” (i.e., dependence) among edges
    - Does work iff edges are independent – i.e., the logistic case
- **Still, useful aid in interpretation**
  - Can consider probability of  $y_{ij}$  under various scenarios to understand *local* model behavior

# Fitting ERGs to Data

- **After positing a model, generally want to estimate parameters from data**
  - Benefit of the framework: standardized inferential framework
- **Most common current method: maximum likelihood estimation**
  - Find  $\theta^*$  that maximizes  $\Pr(\mathbf{Y}=\mathbf{y}_{\text{obs}}|\mathbf{t},\theta^*,\mathcal{Y})$ 
    - Exists for regular model so long as observed data is “non-extreme” (in a specific sense to be discussed later); always unique
    - Also has first order interpretation:  $\mathbf{E}_{\theta^*} \mathbf{t}(\mathbf{Y})=\mathbf{t}(\mathbf{y}_{\text{obs}})$
  - Approximate standard errors based on Hessian of maximized likelihood
    - Not guaranteed, but simulation studies suggest to be reasonable
- **Important point: estimation is conditional on the support**
  - Often taken for granted, but (as we shall see!) can be consequential

# ERG Fitting Using `ergm`

- **Dedicated statnet package for fitting, simulating models in ERG form**
- **Basic call structure: `ergm(y~term1(arg)+term2(arg))`**
  - `y`: network object
  - `term1`, `term2`, etc.: terms to use (see `?"ergm-terms"`)
    - `arg`: where relevant, arguments to the term functions
- **Output: `ergm` object**
  - `Summary`, `print`, and other methods can be used to examine it
  - `simulate` command can also be used to take draws from the fitted model



# ERGM Inference from Partially Observed Data

- What if  $\mathbf{y}_{\text{obs}}$  is not an entire adjacency matrix, but only a subset thereof?
- If pattern of missingness is ignorable, natural to use a "latent missing data approach" (Handcock and Gile, 2007)
  - Treat unknown portion of  $\mathbf{y}_{\text{obs}}$  as latent variables; by assumption, must be drawn from the estimated model
  - Conditional probability of the unknown data is similar to original ERGM, with restricted normalizing factor (sum is over all observable "completions" of  $\mathbf{y}_{\text{obs}}$ ):

$$\Pr(Y^{\bar{o}} = \mathbf{y}^{\bar{o}} | Y^o = \mathbf{y}^o, X, \mathcal{Y}) = \frac{\exp(\theta^T \mathbf{t}(\mathbf{y}^{\bar{o}} \cup \mathbf{y}^o, X))}{\sum_{\mathbf{y}'^{\bar{o}}: \mathbf{y}'^{\bar{o}} \cup \mathbf{y}^o \in \mathcal{Y}} \exp(\theta^T \mathbf{t}(\mathbf{y}'^{\bar{o}} \cup \mathbf{y}^o, X))}$$

# Partially Observed Data, Cont.

- **Observed data likelihood is then proportional to sum of likelihoods for "compatible" realizations of  $y_{\text{obs}}|y^o$ :**

$$\Pr(Y^o = y^o | X, \mathcal{Y}) \propto \frac{\sum_{y'^o: y'^o \cup y^o \in \mathcal{Y}} \exp(\theta^T \mathbf{t}(y'^o \cup y^o, X))}{\sum_{y' \in \mathcal{Y}} \exp(\theta^T \mathbf{t}(y', X))}$$

- Automatically computed by default in `ergm`
- Coefficients, other properties have standard interpretation
- Will lose power, of course! (Perhaps substantially)
- **Reminder: not a fix for biased samples!**
  - If missingness depends on parameters other than through  $X$ ,  $y^o$ , then results may be very misleading
    - E.g., sampling based on unobserved properties of missing nodes