CIDnetworks: How It Works

A.C. Thomas, December 12 2013

The Conditionally Independent Dyadic network framework is built on the notion that, compared to the Exponential Random Graph model, each dyad is independent of the others in a network conditional on whatever we can infer about nodal properties in the network, as well as any covariates on the edges that are observable.

Undirected Networks

In the undirected case, there is exactly one tie per dyad to consider. Suppose we have N total nodes in the system. This means there are $\binom{N}{2}$ potential edges in the network that can be observed and modelled; let i and j index the nodes that are connected for each potential edge. Suppose that the total number we observe is labelled M, though in practice this is typically the full set of $\binom{N}{2}$.

The first underlying generative principle for this network type is that each tie has a latent strength of connectivity Z_{ij}, which is modelled as a Gaussian random variable with its own mean and variance,

$$Z_{ij} \sim N(\mu_{ij}, \sigma^2_{ij}).$$

The form of this mean and variance then determines how this latent strength is derived. We describe how to obtain different forms of μ_{ij} in the subsequent sections; for these examples, we have one variance, $\sigma^2_{ij} = \sigma^2$.

This form is used partly because of its computational advantage, but also because of its flexibility. If we have a network whose edges are Gaussian in nature, we simply set $Y_{ij} = Z_{ij}$.

If we want a network with binary edges – as the vast majority of observed networks are said to be – then we use the standard generative mechanism for the probit model to set our outcome of zero or one,

$$Y_{ij} = \mathbb{I}(Z_{ij} > 0),$$

so that the probability of an edge existing depends directly on the latent tie strength and indirectly on the mean of that tie μ_{ij}. In this case we fix $\sigma^2 = 1$ to identify the model.

We can also extend this to any ordinal data outcomes by considering additional hurdles. For example, consider an ordinal outcome that can take values of $(0, 1, 2, 3)$. With the latent continuous variable Z_{ij}, we can generate an ordinal variable by taking three cutoff values and determining the latent value with respect to these cutoffs. We already have one in the pre-existing zero; let the others be c_1 and c_2 and have them both be greater than zero. The

outcome is then obtained as

$$Y_{ij} = 0 * \mathbb{I}(Z_{ij} < 0) + 1 * \mathbb{I}(0 < Z_{ij} < c_1) + 2 * \mathbb{I}(c_1 < Z_{ij} < c_2) + 3 * \mathbb{I}(c_2 < Z_{ij}).$$

Given that the latent strength Z_{ij} determines the value of the tie itself, assume for now that this value is known in inference tasks.

Computational Mechanisms

We embrace a fully Bayesian interpretation of this mechanism, mainly because it provides us with a convenient computational framework for estimating network parameters. We use Markov Chain Monte Carlo for this purpose, using a cyclic Gibbs sampler on each parameter with direct draws when possible and Metropolis proposals when not.

Specifying μ_{ij}

By specifying the latent strength of a tie first, we have the opportunity to make it an additive collection of terms each corresponding to a different type of dependence structure. First, we specify each of a number of different forms of dependence, as well as each of the mechanisms by which we can sample from their distributions.

First, we specify that the latent strengths share a common grand mean μ, which has a prior distribution

$$\mu \sim N(\mu_0, \sigma_\mu^2).$$

Specifying each latent strength as $Z_{ij} = \mu + \mu'_{ij} + \varepsilon_{ij}$, where $\varepsilon_{ij} \sim N(0, \sigma^2)$, we then have

$$Z_{ij} - \mu'_{ij} = \mu + \varepsilon_{ij},$$

so that we can obtain a direct draw of μ, conditional on the means of each of the M observed potential edges, from the standard normal distribution form

$$(\mu | Z, \mu', \sigma^2) \sim N\left(\left(\frac{M}{\sigma^2} + \frac{1}{\sigma_\mu^2}\right)^{-1}\left(\frac{\sum_{i,j}(Z_{ij} - \mu'_{ij})}{\sigma^2} + \frac{\mu_0}{\sigma_\mu^2}\right), \left(\frac{M}{\sigma^2} + \frac{1}{\sigma_\mu^2}\right)^{-1}\right).$$

Similarly for σ^2, we can specify a semi-conjugate prior distribution $InverseGamma(a_{\sigma^2}, b_{\sigma^2})$, and with the likelihood

$$Z_{ij} - \mu'_{ij} - \mu = \varepsilon_{ij},$$

we have the standard semi-conjugate draw for the variance parameter,

$$(\sigma^2|\mu,\mu_{ij},Z) \sim InverseGamma\left(a_{\sigma^2}+\frac{M}{2}\,,b_{\sigma^2}+\frac{\sum_{i,j}(Z_{ij}-\mu'_{ij}-\mu)^2}{2}\right).$$

Covariates on Edges (COV)

For a covariate known on each potential edge on a network, this model is functionally identical to the standard linear model: each edge has a vector of k covariates and a corresponding k-vector of coefficients labelled β:

$$Z_{ij} = \mu + X_{ij}\beta + \varepsilon_{ij}.$$

With the transformation $Z_{ij} - \mu = X_{ij}\beta + \varepsilon_{ij}$, we have the standard linear model expression. With prior distribution $\beta \sim N_k(\beta_0, \Sigma_\beta)$, and with X as an M-by-k matrix, we have

$$(\beta|\mu,\sigma^2,Z) \sim N_k\left(\left(X'X/\sigma^2+\Sigma_\beta^{-1}\right)^{-1}\left(X'Z/\sigma^2+\Sigma_\beta^{-1}\beta_0\right),\left(X'X/\sigma^2+\Sigma_\beta^{-1}\right)^{-1}\right)$$

Taking each item in sequence, here is the order for our Gibbs sampler, given that the unspecified terms are conditioned on at each step:

1. Draw μ from its semiconjugate draw.
2. Draw σ^2 from its semiconjugate draw, if necessary.
3. Draw β as specified above.

Sender-Receiver Intercepts (SR)

Each node has an associated strength that is above or below the average of their group. These can either function as “fixed” or “random” effects on these outcomes,

$$Z_{ij} = \mu + \alpha_i + \alpha_j + \varepsilon_{ij},$$

with a prior distribution on each α,

$$\alpha_i \sim N(0, \sigma_\alpha^2).$$

These effects are fixed if σ_α^2 is a constant and partially pooled if we impose a prior distribution, such as $InverseGamma(a_\alpha, b_\alpha)$; this is the form that is enabled by default in CIDnetworks. Let Σ_α be the n-by-n diagonal matrix whose non-zero terms are each σ_α^2.

If σ_α^2 is held constant, by definition or conditioning, then $\alpha_i + \alpha_j$ defines a sparse M-by-n design covariance matrix Q that multiplies the full vector of effects α; each row has exactly two entries equal to one and is of full column rank.

The steps for one iteration of the Gibbs sampler:

1. Draw μ from its semiconjugate draw.
2. Draw σ^2 from its semiconjugate draw, if necessary.
3. Draw α from

$$(\alpha|\mu,\sigma^2,\sigma_\alpha^2,Z) \sim N_k\left(\left(Q'Q/\sigma^2+\Sigma_\alpha^{-1}\right)^{-1}\left(Q'(Z-\mu)/\sigma^2\right),\left(Q'Q/\sigma^2+\Sigma_\alpha^{-1}\right)^{-}\right.$$

4. Draw σ_α^2 from

$$(\sigma_\alpha^2|\alpha,\mu,Z) \sim InverseGamma\left(a_\alpha+\frac{n}{2}\,,b_\alpha+\frac{\sum_i \alpha_i^2}{2}\right).$$

Latent Space Model (LSM)

Each node i is assigned a position d_i in a latent space (typically Euclidean). The latent strength of the tie decreases as the distance between points increases,

$$Z_{ij} = \mu - |d_i - d_j| + \varepsilon_{ij}.$$

This function has the downside of lacking a closed-form solution for a complete conditional distribution, which is balanced by the usefulness of the latent space position as a descriptor.

The inverted version of this model has the strength of tie increase with distance ($\mu + |d_i - d_j|$), though this does not have the advantage of interpretability. It is computed with the same algorithm.

The steps for one iteration of the Gibbs sampler:

1. Draw μ from its semiconjugate draw.
2. Draw σ^2 from its semiconjugate draw, if necessary.
3. For each $i \in \{1,\ldots,n\}$, propose a new d_i in the latent space by some proposal distribution (in this case, a multivariate normal centered at the original d_i). Accept with probability equal to the standard Metropolis ratio.

Latent Vector Model (LVM)

Each node i is assigned a vector q_i in a latent space (typically Euclidean). The latent strength of the tie is a direct function of the inner product between the two vectors:

$$Z_{ij} = \mu + {q_i}'q_j + \varepsilon_{ij}.$$

If $i \neq j$ for all edges in the system, then if we condition on all node effects except i, then the likelihood is identical to a linear model for all edges with node i and constant for all others. Let each q_i have dimension k and prior distribution $q_i \sim N_k(0,\Sigma_q)$

The steps for one iteration of the Gibbs sampler:

1. Draw μ from its semiconjugate draw.
2. Draw σ^2 from its semiconjugate draw, if necessary.
3. For each $i \in \{1, \ldots, n\}$, select the relevant edges. Construct R_i, a matrix whose rows are the latent vectors for the nodes that comprise potential edges with i, and Z_i, the outcomes corresponding to those rows. Then draw from the conditional distribution

$$(q_i|\mu, \sigma^2, Z) \sim N_k\left(\left({R_i}'R_i/\sigma^2 + \Sigma_q^{-1}\right)^{-1}\left({R_i}'Z_i/\sigma^2\right), \left({R_i}'R_i/\sigma^2 + \Sigma_q^{-1}\right)^{-1}\right)$$

Stochastic Block Model (SBM)

Each node belongs to one of k discrete blocks. All edges between members of one block and another (possibly the same block) have the same mean value.

Let S_i be a k-vector with a 1 in the position corresponding to the block membership and 0 otherwise. Let B be a k-by-k symmetric matrix of mean values, so that B_{ab} is the mean value if i and j belong to blocks a and b respectively. The outcome is then generated as

$$Z_{ij} = \mu + S_i B S_j + \varepsilon_{ij}.$$

Each membership vector S_i has prior probability $1/k$ of belonging to each group. Each element of the matrix B has prior distribution $N(0, \sigma_b^2)$.

The steps for one iteration of the Gibbs sampler:

1. Draw μ from its semiconjugate draw.
2. Draw σ^2 from its semiconjugate draw, if necessary.
3. For each node i, compute the resulting likelihood of the data for its membership in each of the k groups. Choose one of these groups with probability proportional to the likelihood (a direct draw from the discrete distribution).
4. For each entry of the block matrix B_{ab}, find all edges whose group memberships match that entry (totalling M_{ab}. Conditional on the group memberships, this is now a simple semi-conjugate normal draw:

$$(B_{ab}|\mu, Z, S) \sim N\left(\left(\frac{M_{ab}}{\sigma^2} + \frac{1}{\sigma_b^2}\right)^{-1}\left(\frac{\sum_{i,j \in a,b}(Z_{ij} - \mu)}{\sigma^2}\right), \left(\frac{M_{ab}}{\sigma^2} + \frac{1}{\sigma_b^2}\right)^{-1}\right).$$

Mixed-Membership Stochastic Block Model (MMSBM)

In the SBM, each node belongs strictly to one block. In the mixed membership case, the membership of a node can change with respect to each partner with which it can share an edge, with respect to an underlying distribution of block membership probabilities. The underlying mean is now

$$Z_{ij} = \mu + S_{ij}BS_{ji} + \varepsilon_{ij}.$$

Each partial membership vector S_{ij} derives from a nodal mixed membership vector π_i,

$$S_{ij} \sim Multinomial(1, \pi_i),$$

and each π_i has a prior Dirichlet distribution $Dir(n_o(\frac{1}{k}, \ldots, \frac{1}{k}))$. Again, each element of the matrix B has prior distribution $N(0, \sigma_b^2)$.

The steps for one iteration of the Gibbs sampler:

1. Draw μ from its semiconjugate draw.
2. Draw σ^2 from its semiconjugate draw, if necessary.
3. For each directed node pair (i, j), compute the resulting likelihood of the data for its membership in each of the k groups. Choose one of these groups with probability proportional to the likelihood times the prior (π_i) (a direct draw from the discrete distribution)
4. For each mixed membership vector π_i, we have multinomial data corresponding to the number of partial memberships in each block, and a Dirichlet prior, yielding a direct draw from the Dirichlet posterior.
5. For each entry of the block matrix B_{ab}, find all edges whose partial group memberships match that entry (totalling M_{ab}. Conditional on the group memberships, this is again a simple semi-conjugate normal draw:

$$(B_{ab}|\mu, Z, S) \sim N\left(\left(\frac{M_{ab}}{\sigma^2} + \frac{1}{\sigma_b^2}\right)^{-1}\left(\frac{\sum_{i,j\in a,b}(Z_{ij} - \mu)}{\sigma^2}\right), \left(\frac{M_{ab}}{\sigma^2} + \frac{1}{\sigma_b^2}\right)^{-1}\right).$$

Hierarchical Block Model (HBM)

There is a tree with k internal nodes labelled $m \in \{1, \ldots, k\}$, wherein each internal node has mean value C_m. Each node i on the network belongs to one of these internal nodes, kept as vector S_i; the mean value of Z_{ij} is the value of the closest common ancestor to both i and j.

$$Z_{ij} = \mu + C_{common(i,j)} + \varepsilon_{ij}$$

All nodes have equal prior probability of belonging to an internal node.

(more to describe here to tell the story.)

The steps for one iteration of the Gibbs sampler:

1. Draw μ from its semiconjugate draw.
2. Draw σ^2 from its semiconjugate draw, if necessary.
3. For each node i, compute the resulting likelihood of the data for its membership in each of the k groups. Choose one of these groups with probability proportional to the likelihood (a direct draw from the discrete distribution). If its departure would leave fewer than two

nodes attached to an internal node, propose a Metropolis step in which it switches with another node and accept with the standard Metropolis ratio.

4. For each entry of the block matrix C_m, find all edges whose common ancestor corresponds to that internal node, totalling M_m. Conditional on the group memberships, this is now a simple semi-conjugate normal draw:

$$(C_m|\mu, Z, S) \sim N\left(\left(\frac{M_m}{\sigma^2} + \frac{1}{\sigma_b^2}\right)^{-1}\left(\frac{\sum_{i,j\in m}(Z_{ij} - \mu)}{\sigma^2}\right), \left(\frac{M_m}{\sigma^2} + \frac{1}{\sigma_b^2}\right)^{-1}\right).$$

Post-processing

We have two basic classes of model with latent information: continuous and discrete position. For each we have a standard way of transforming the data for easy analysis, in a way that simply identifies the model rather than changing it in a meaningful way.

Continuous latent space/vector values

After every iteration of the Gibbs sampler, we perform the following realignment:

1. Recenter the latent positions to have mean zero.
2. Rotate the positions so that node 1 lies on the positive x-axis.
3. Rotate the positions so that node 2 lies in the x-y plane with positive y value.
4. Repeat rotations so that node i lies in the i-dimensional hyperplane with positive i-dimensional value, up to the dimension of the space itself.

This routine specifies a unique alignment of points that makes it easier to compare the stability of the model fit, particularly when the number of nodes greatly exceeds the dimension.

Discrete block memberships

After every iteration of the Gibbs sampler, we perform the following realignment:

1. Permute the block numbers so that block 1 contains either node 1 or the maximum probability for node 1.
2. Continue permutations so that node i is contained in block i, unless it is already in a lower-numbered block.
3. Conclude the permutations so that lower-numbered nodes have the smallest block numbers possible.

Like in the latent space case, this routine specifies a unique alignment of points that makes it easier to compare the stability of the model fit, particularly when the number of nodes greatly exceeds the dimension of the block.

Multiple Contributions to μ_{ij}

Each of the previous components can be included by themselves, or in combination with others, as a generative network model. The Gibbs sampler construction and additive form makes it easy to sample from each piece in sequence.

Consider a model with three subclass components in addition to the grand intercept: a covariate piece, a latent space, and a stochastic block model. Suppose this is also a binary network, so that once all the components are specified, the generative model is:

$$Z_{ij} = \mu + X_{ij}\beta - |d_i - d_j| + S_i B S_j + \varepsilon_{ij}$$

$$Y_{ij} = \mathbb{I}(Z_{ij} > 0)$$

Note that since this is the binary case, $Var\varepsilon_{ij} = 1$. The full Gibbs sequence for one iteration cycles over each of the sub-components:

1. Collect and sample for μ, noting that

$$\left(Z_{ij} - X_{ij}\beta + |d_i - d_j| - S_i B S_j\right) = \mu + \varepsilon_{ij}.$$

2. Collect and sample for β:

$$\left(Z_{ij} - \mu + |d_i - d_j| - S_i B S_j\right) = X_{ij}\beta + \varepsilon_{ij}.$$

3. Collect and sample for the latent space positions $|d_i - d_j|$:

$$\left(Z_{ij} - \mu - X_{ij}\beta - S_i B S_j\right) = -|d_i - d_j| + \varepsilon_{ij}$$

4. Collect and sample for the stochastic block parameters $S_i B S_j$:

$$\left(Z_{ij} - \mu - X_{ij}\beta + |d_i - d_j|\right) = S_i B S_j + \varepsilon_{ij}$$

5. Draw new latent variables Z_{ij} from their truncated normal distributions, depending on the value of each Y_{ij}.

For $Y_{ij} = 0$:

$$Z_{ij} \sim TruncatedNormal(\mu + X_{ij}\beta - |d_i - d_j| + S_i B S_j, 1, upper = 0)$$

For $Y_{ij} = 1$:

$$Z_{ij} \sim TruncatedNormal(\mu + X_{ij}\beta - |d_i - d_j| + S_i B S_j, 1, lower = 0)$$