Supplementary Materials: Visual Inference for a Social Network Model

Blah

A Model Details

The continuous-time Markov chain (CTMC) family of models we use were first introduced in Snijders (1996). We describe their basic structure here, and full detail can be found in Snijders (1996); Snijders et al. (2010).

Rate Function: The rate function dictates when network changes are made and which actor can make them. An actor, i, is chosen make a change in its ties one of the other nodes j. In general, the rate function can include structural and node covariate parameters into account so that each actor has a different rate of change. However, we choose a simple rate function that is constant over all nodes in a given time period, because we focus on interpreting the parameters of the objective function which directly impact the overall network structure. We denote the rate from t_m to t_{m+1} as α_m for $m = 1, \ldots, M-1$, where M is the number of time points at which the network was observed. Using this notation, the *waiting time* to the next chance for actor i to make a change is exponentially distributed with expected value α_m^{-1}. Since the rate is the same for all actors, the waiting time for *any* actor to get the opportunity to change its set of ties is also exponentially distributed with expected value $(n\alpha_m)^{-1}$.

Objective Function: After actor i is selected to make a change, it randomly picks one of its current ties, x_{ij}, to change. Actor i aims to maximize the objective function f_i given the current state of the network, x and the node-level covariates, $\mathbf{Z}$. This function is defined as:

$$f_i(x, \boldsymbol{\beta}, \mathbf{Z}) = \sum_{k=1}^{K} \beta_k s_{ik}(x, \mathbf{Z}), \tag{1}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)$ are additional model parameters, each associated with some statistics, $s_{i1}(x, \mathbf{Z}), \ldots, s_{iK}(x, \mathbf{Z})$, calculated for actor i at the current network state x. At least two parameters must be included in the objective function: density and reciprocity (Ripley et al., 2017). We denote the density, or out-degree, parameter by β_1 and the associated statistic as $s_{i1}(x) = \sum_j x_{ij}$ and we denote the reciprocity parameter by β_2 and the associated statistic as $s_{i2}(x) = \sum_j x_{ij}x_{ji}$. We will refer to the very simple model with only these two parameters in the objective function as model M1. We define additional parameters and models of interest in Section **??** Version 1.2-3 of `RSiena` (Ripley et al., 2013), the software we use to fit CTMC models to data, provides over 80 possible effects that can be included in the objective function.

The objective function $f_i(x, \boldsymbol{\beta}, \mathbf{Z})$ dictates the *transition probability*, p_{ij} of the network changing from its current state x to the state $x(i \rightsquigarrow j)$, which is identical to x except for x_{ij}: $x_{ij}(i \rightsquigarrow j) = 1 - x_{ij}$. The transition probability is

$$p_{ij} = \frac{\exp\{f_i(x(i \rightsquigarrow j), \boldsymbol{\beta}, \mathbf{Z})\}}{\sum_h \exp\{f_i(x(i \rightsquigarrow h), \boldsymbol{\beta}, \mathbf{Z})\}}, \tag{2}$$

dictating which edge node i changes. Thus, the actor is more likely to make changes that increase the value of their objective function, and no change is most likely when any change decreases the value of the objective function.

A.1 Goodness-of-Fit Testing

The software `RSiena` contains methods for performing goodness-of-fit tests for the CTMC models. The `sienaGOF()` function performs goodness-of-fit testing as follows:

1. Auxiliary statistics, such as the cumulative outdegree distribution on the nodes, are computed on the observed data ($\mathbf{u}_d$) and on N observations simulated from the model ($\mathbf{u}_1 \ldots \mathbf{u}_N$).

2. The mean $\overline{\mathbf{u}}$ and covariance matrix $\mathbf{S}$ are computed from the N simulations, and the Mahalanobis distance, $d_M(\mathbf{u}_d)$ from the observed statistics to the distribution of the simulated statistics is computed:

$$d_M(\mathbf{u}_d) = \sqrt{(\mathbf{u}_d - \overline{\mathbf{u}})'\mathbf{S}^{-1}(\mathbf{u}_d - \overline{\mathbf{u}})} \tag{3}$$

3. The Mahalanobis distance for each of the N simulations is calculated and $d_M(\mathbf{u}_d)$ is compared to this distribution of distances.

4. An empirical p-value is found by computing the proportion of simulated distances found in step 4 that are as large or larger than $d_M(\mathbf{u}_d)$.

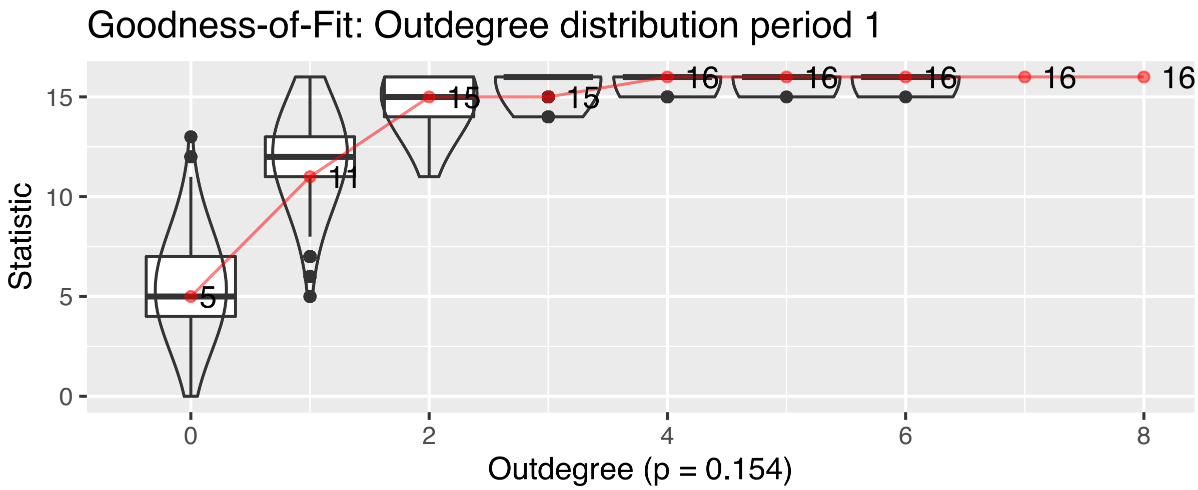

Figure 1: An example of what a goodness-of-fit plot from `RSiena` looks like. The overlaid boxplots and violin plots show the distribution of each of the outdegree count values on the simulated networks, and the red points and lines are the observed data values.

B Data

Details of how this data can be downloaded are provided by François Briatte at `github.com/briatte/congress`. In the US Senate, senators often show support for a piece of legislation by co-sponsoring a bill authored by one of their colleagues. In a co-sponsorship network, ties are directed from senator i to senator j when senator i signs on as a co-sponsor to the bill that senator j authored. There are many hundreds of ties between senators when they are connected in this way, so we simplify the network by computing a single value for each senator-senator collaboration called the *weighted propensity to co-sponsor* (WPC). This value is defined in Gross et al. (2008) as

$$WPC_{ij} = \sum_{b=1}^{B_j} \frac{Y_{ij(b)}}{c_{j(b)}} \left(\sum_{b=1}^{B_j} \frac{1}{c_{j(b)}} \right)^{-1} \tag{4}$$

where B_j is the number of bills in a congressional session authored by senator j, $c_{j(b)}$ is the number of co-sponsors on senator j's b^{th} bill, where $b \in \{1, \ldots, B_j\}$, and $Y_{ij(b)}$ is an indicator variable that senator i co-sponsored senator j's b^{th} bill. This measure ranges in value from 0 to 1, where $WPC_{ij} = 1$ if senator i is a co-sponsor on every one of senator j's bills and $WPC_{ij} = 0$ if senator i is never a co-sponsor any of senator j's bills. To simplify the problem, we construct a network with binary edges as follows:

$$x_{ij} = \begin{cases} 1 & WPC_{ij} > 0.25 \\ 0 & WPC_{ij} \leq 0.25 \end{cases} \tag{5}$$

so that only strong ties between senators are in the network.

C Results

C.1 Visual Power

References

Gross, J. H., Kirkland, J. H., and Shalizi, C. R. (2008), "Cosponsorship in the U.S. Senate: A Multilevel Two-Mode Approach to Detecting Subtle Social Predictors of Legislative Support," *Unpublished Manuscript*.

Ripley, R., Boitmanis, K., and Snijders, T. A. (2013), *RSiena: Siena - Simulation Investigation for Empirical Network Analysis*, r packa ge version 1.1-232.

Ripley, R. M., Snijders, T. A., Boda, Z., Vörös, A., and Preciado, P. (2017), "Manual for RSiena," Tech. rep., `https://www.stats.ox.ac.uk/~snijders/siena/RSiena_Manual.pdf`.

Snijders, T. A., van de Bunt, G. G., and Steglich, C. E. (2010), "Introduction to stochastic actor-based models for network dynamics," *Social Networks*, 32, 44 – 60, dynamics of Social Networks.

Snijders, T. A. B. (1996), "Stochastic actor-oriented models for network change," *Journal of Mathematical Sociology*, 21, 149–172.

Parameter	Estimate	Std Error	p-value	Odds Multiplier
η_{1+}	37.297	6.573	<0.0001‡	$> 10^4$
η_{1-}	-9.933	3.473	0.0042†	$< 10^{-4}$
γ_{1+}	75.356	13.532	<0.0001‡	$> 10^4$
γ_{1-}	-14.504	4.804	0.0025†	$< 10^{-4}$
η_{2+}	-6.833	4.466	0.1260	0.0011
η_{2-}	-17.001	2.236	<0.0001‡	$< 10^{-4}$
γ_{2+}	13.771	7.446	0.0644*	$> 10^4$
γ_{2-}	-229.16	31.306	<0.0001‡	$< 10^{-4}$
η_{3+}	-2.801	0.949	0.0032†	0.0608
η_{3-}	-2.644	0.811	0.0011†	0.0711
γ_{3+}	4.474	1.389	0.0013†	87.7507
γ_{3-}	-1.108	0.609	0.0690*	0.3304
η_{4+}	-2.078	0.954	0.0293**	0.1252
η_{4-}	-2.692	1.322	0.0417**	0.0678
γ_{4+}	4.247	2.147	0.0479**	69.8675
γ_{4-}	2.403	2.187	0.2719	11.0585
η_{5+}	-5.84	2.989	0.0507*	0.0029
η_{5-}	-4.686	0.86	<0.0001‡	0.0092
γ_{5+}	3.264	1.756	0.0630*	26.1487
γ_{5-}	-2.176	0.387	<0.0001‡	0.1136
η_{6+}	-1.164	1.226	0.3425	0.3123
η_{6-}	-5.929	1.28	<0.0001‡	0.0027
γ_{6+}	5.76	3.524	0.1021	317.4753
γ_{6-}	15.092	3.605	<0.0001‡	$> 10^4$
σ^2_{δ}	0.564	–	–	–
σ^2_{ϵ}	0.342	–	–	–

Table 1: Summary of the results from fitting the model given in Equation **??**. Significance levels: * - < 0.10; ** - < 0.05; † - < 0.01; ‡ - < 0.001