SMOGS: Social Network Metrics of Game Success Supplementary Material

Fan Bu **Sonia Xu** **Katherine Heller** **Alexander Volfovsky**

Department of Statistical Science, Duke University

A DERIVATION FOR THE PARAMETER INFERENCE ALGORITHM

Here we lay out the details of the full conditional distributions used in step 1 and step 2 in the sampling algorithm in Section 3.2.

Since not all the players are on the court all the time in a basketball game, parameters relevant to a player are inferred using only the observations at the moments when he was playing.

A.1 Full Conditionals for the Covariate Coefficients (Step 1)

For each player pair (i, j), let $\theta_{i,j}$ be the vector of all log-risks player i passing to j across all the games in the data. For each entry $\theta_{i,j,g}(t)$ in the vector (suppose that the corresponding game is g), let

$$\tilde{\theta}_{i,j,g}(t) = \theta_{i,j,g}(t) - u_{i,g}^T v_{j,g}, \tag{1}$$

and we have

$$\tilde{\theta}_{i,j,g}(t) = X_{i,j,g}(t)^T \beta_{i,j} + \epsilon_{i,j,g}(t). \tag{2}$$

Re-writing the above in matrix form (by stacking together all the entries across all the games), we get

$$\tilde{\theta}_{i,j} = \mathbf{X}_{i,j}^T \beta_{i,j} + \epsilon_{i,j}. \tag{3}$$

Then under the Gaussian model for independent errors $\epsilon_{i,j,g}(t)$ and a normal prior for $\beta_{i,j}$,

$$\begin{aligned} \epsilon_{i,j,g}(t) &\sim N(0, 1), \\ \beta_{i,j} &\sim N(b_{i,j}, \tau^2 I_7), \end{aligned} \tag{4}$$

$p(\beta_{i,j}|\Theta, \mathbf{U}, \mathbf{V})$ is a multivariate normal distribution,

$$\beta_{i,j}|\Theta, \mathbf{U}, \mathbf{V} \sim N(m_{i,j}, M_{i,j}), \tag{5}$$

where

$$\begin{aligned} M_{i,j} &= (I_7/\tau^2 + \mathbf{X}_{i,j}^T \mathbf{X}_{i,j})^{-1}, \\ m_{i,j} &= M_{i,j}(b_{i,j}/\tau^2 + \mathbf{X}_{i,j}^T \tilde{\theta}_{i,j}). \end{aligned} \tag{6}$$

In Eq (4), we adopt a diffuse prior for $\beta_{i,j}$, with $\tau = 1000$ and b_{ij} as the OLS estimate for $\beta_{i,j}$ assuming $U = V = 0$.

A.2 Full Conditionals for the Multiplicative Latent Effects (Step 2)

For each game g and each player i who played in game g, let $TM_g(i)$ be the set of players who shared the court with i in game g. For $j \in TM_g(i)$ and time t at which i possessed the ball and j was on the court, we have

$$\bar{\theta}_{i,j,g}(t) = u_{i,g}^T v_{j,g} + \epsilon_{i,j,g}(t), \tag{7}$$

where

$$\bar{\theta}_{i,j,g}(t) = \theta_{i,j,g}(t) - X_{i,j,g}(t)^T \beta_{i,j}. \tag{8}$$

Re-writing Eq (7) in matrix form, we have

$$\bar{\theta}_{i,g} = V_{TM_g(i),g}^T u_{i,g} + \epsilon_{i,g}, \tag{9}$$

where $V_{TM_g(i),g}$ is the matrix with each row being a receiver-specific latent effect vector $v_{j,g}$ in Eq (7).

Under a normal prior $u_{i,g} \sim N(0, I_R)$, $p(U_g[i,]|\Theta, \mathbf{V}, \beta)$ is a multivariate normal distribution,

$$U_g[i,]|\Theta, \mathbf{V}, \beta \sim N(w_{i,g}, W_{i,g}), \tag{10}$$

where

$$\begin{aligned} W_{i,g} &= (I_R + V_{TM_g(i),g}^T V_{TM_g(i),g})^{-1}, \\ w_{i,g} &= W_{i,g}(I_R + V_{TM_g(i),g}^T \bar{\theta}_{i,g}). \end{aligned} \tag{11}$$

Very similarly, for each game g and each player j who played in game g, let $PO_g(j)$ be the set of players who shared the court with j and ever possessed the ball in game g. For $i \in PO_g(j)$ and time t at which i was the ballcarrier and j was on the court, we have (with a slight abuse of notation)

$$\bar{\theta}_{i,j,g}(t) = u_{i,g}^T v_{j,g} + \epsilon_{i,j,g}(t). \tag{12}$$

Re-writing Eq (12) in matrix form, we have

$$\bar{\theta}_{j,g} = U^T_{PO_g(j),g} v_{j,g} + \epsilon_{j,g}, \tag{13}$$

where $U_{PO_g(j),g}$ is the matrix with each row being a sender-specific latent effect vector $u_{i,g}$ in Eq (12).

Under a normal prior $v_{j,g} \sim N(0, I_R)$, $p(V_g[j,]|\Theta, \mathbf{U}, \beta)$ is a multivariate normal distribution,

$$V_g[j,]|\Theta, \mathbf{U}, \beta \sim N(w'_{j,g}, W'_{j,g}), \tag{14}$$

where

$$\begin{aligned} W'_{j,g} &= (I_R + U^T_{PO_g(j),g} U_{PO_g(j),g})^{-1}, \\ w'_{j,g} &= W'_{j,g}(I_R + U^T_{PO_g(j),g} \bar{\theta}_{j,g}). \end{aligned} \tag{15}$$

B ADDITIONAL PLOTS OF MULTIPLICATIVE LATENT EFFECTS

B.1 Additional Plots for the AME Model in Section 2

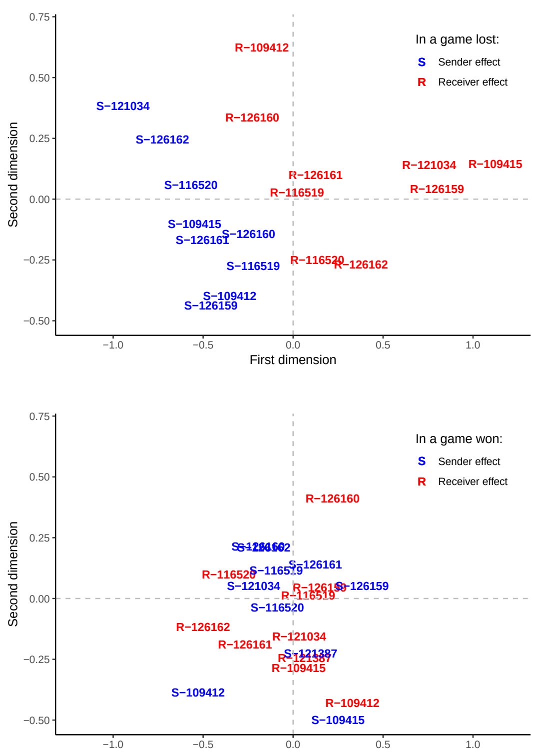


Figure 1: Learned multiplicative sender-specific effects (in blue) and receiver-specific effects (in red) by the AME model in a lost game (top) and a won game (bottom). Each latent effect vector corresponds to a two-dimensional coordinate represented by a player's id code.

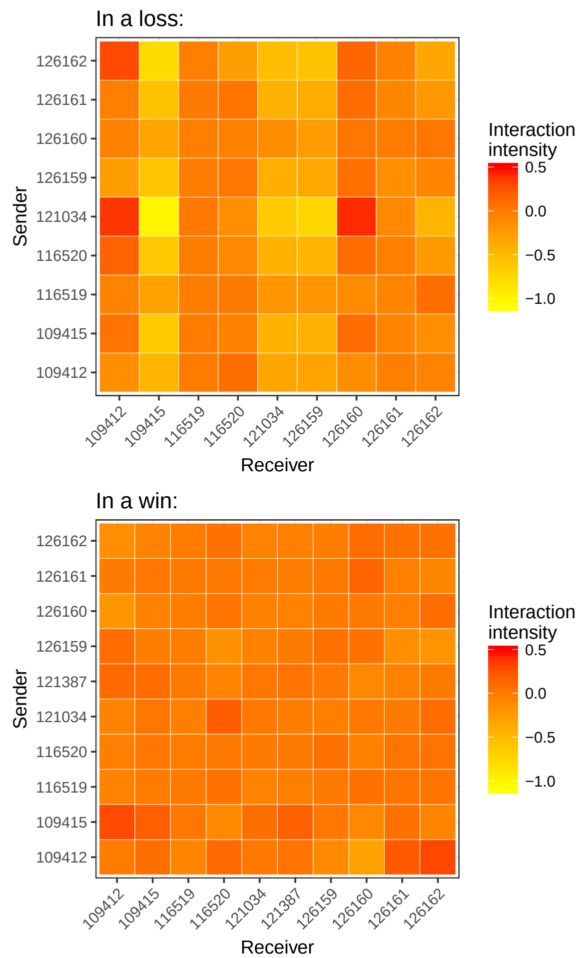


Figure 2: Products of sender- and receiver-specific effects learned in the AME model for all player pairs in a loss (top) and in a win (bottom). Darker color indicates higher frequency of passes. The level of interaction between teammates is significantly higher in a win than in a loss. Furthermore, there are significant passing behavior anomalies in the lost game: player 121034 strongly favors 109412 and 126160 as ball receivers but ignores 109415 as a potential receiver, which is not the case in the victory where his passing choices are more balanced.

B.2 Additional Plots for the Real Data Experiments in Section 4

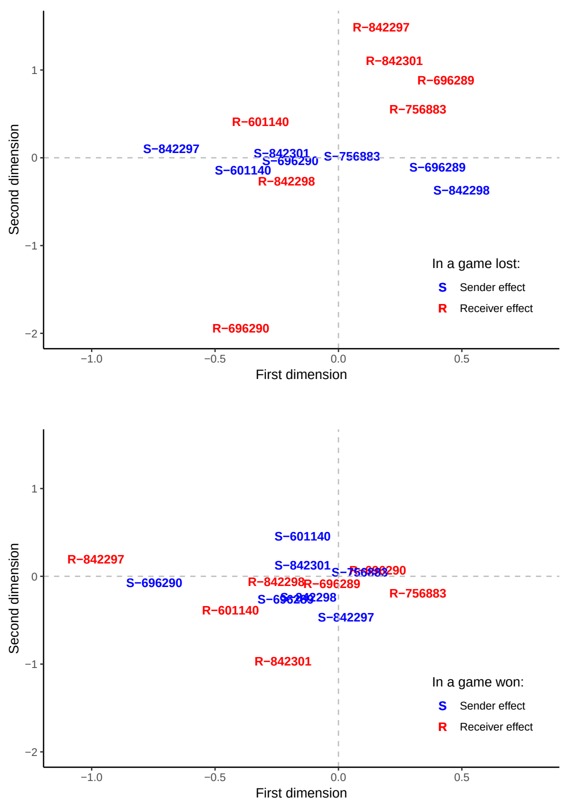
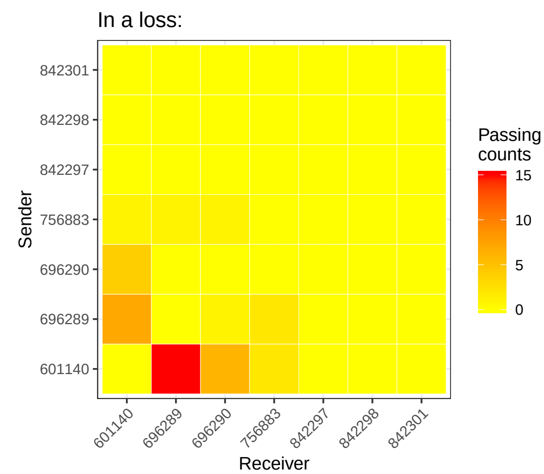
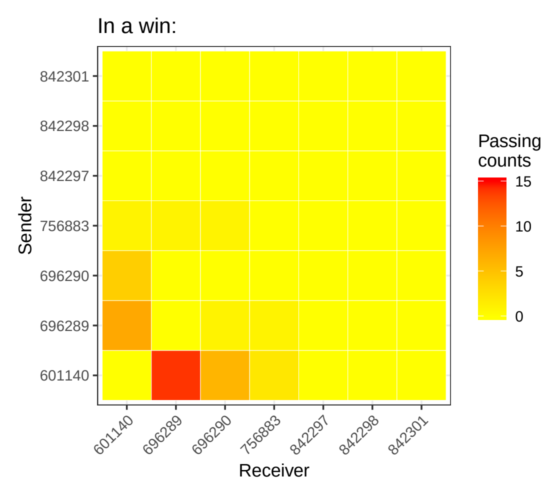


Figure 3: Passing decision multiplicative latent factors in a loss (top) versus in a victory (bottom). Sender-specific effects are marked with "S" plus player id codes in blue, and receiver-specific effects are marked with "R" plus player id codes in red.



(a) Number of passes between players in a lost game.



(b) Number of passes between players in a successful game.

Figure 4: Raw passing counts between home team players in a loss versus in a victory. Darker color indicates more passes made from a sender to a receiver. It is obvious that these two plots are disparate from the plots in Fig.5 in the main text, and that the differences between a win and a loss are much harder to observe if only the raw passing counts are examined.