Appendix

A. 2nd set of analytical solutions

For homophily rewiring, as before let the network exposure term be X, let the prior term be Z, let the outcome be Y, and let the average out-degree for each node be n. If we rewire (1-p)*100% of the ties to perfectly homophilous others, then the new exposure term for an actor i would become

$$X_i' = \frac{n(1-p)Z_i + \sum_{j=1}^{np} Z_j}{n},$$

where Z_j represents the behavior of actor i's original network neighbors, Z_i represents the behavior of actor i's new network neighbors (who hold exactly the same behavior as actor i).

Since $Var(X) = Var(\frac{\sum^{n} Z}{n}) = \frac{Var(Z)}{n}$, we obtain $Var(\frac{\sum^{np} Z}{n}) = \frac{p}{n}Var(Z)$. Also,

$Cov(Z, X') = (1-p)Cov(Z, Z) + pCov(Z, X)$, and

$Cov(X, Z) = sd(Z)sd(X)r_{XZ} = \frac{Var(Z)}{\sqrt{n}} r_{XZ}$.

As a result, the new correlation between network exposure and the prior, after rewiring becomes

$$r_{XZ}^{*} = \frac{\mathrm{cov}(Z, X')}{sd(Z)sd(X')} = \frac{(1-p)\mathrm{cov}(Z,Z) + p\,\mathrm{cov}(Z,X)}{sd(Z)sd(X')}$$

$$= \frac{(1-p)\mathrm{cov}(Z,Z) + p\,\mathrm{cov}(Z,X)}{sd(Z)\sqrt{(1-p)^2 Var(Z) + \frac{p}{n}Var(Z) + 2p(1-p)Cov(X,Z)}}$$

$$= \frac{(1-p)\mathrm{cov}(Z,Z) + p\,\mathrm{cov}(Z,X)}{sd(Z)\sqrt{(1-p)^2 Var(Z) + \frac{p}{n}Var(Z) + \frac{2p(1-p)r_{XZ}}{\sqrt{n}}Var(Z)}}$$

$$= \frac{(1-p)\mathrm{cov}(Z,Z)}{sd(Z)sd(Z)\sqrt{(1-p)^2 + \frac{p}{n} + \frac{2p(1-p)r_{XZ}}{\sqrt{n}}}}$$

$$+ \frac{p\,\mathrm{cov}(Z,X)}{sd(Z)\sqrt{n(1-p)^2 Var(X) + pVar(X) + \sqrt{n}2p(1-p)r_{XZ}Var(X)}}$$

$$= \frac{(1-p)}{\sqrt{(1-p)^2 + \frac{p}{n} + \frac{2p(1-p)r_{XZ}}{\sqrt{n}}}} + \frac{pr_{XZ}}{\sqrt{n(1-p)^2 + p + \sqrt{n}2p(1-p)r_{XZ}}}$$

Similarly, the new correlation between network exposure and the outcome after rewiring becomes

$$r^*_{XY} = \frac{\mathrm{cov}(Y, X')}{sd(Y)sd(X')} = \frac{(1-p)\mathrm{cov}(Y,Z) + p\,\mathrm{cov}(Y,X)}{sd(Y)sd(X')}$$

$$= \frac{(1-p)\mathrm{cov}(Y,Z) + p\,\mathrm{cov}(Y,X)}{sd(Y)\sqrt{(1-p)^2 Var(Z) + \frac{p}{n}Var(Z) + 2p(1-p)Cov(X,Z)}}$$

$$= \frac{(1-p)\mathrm{cov}(Y,Z) + p\,\mathrm{cov}(Y,X)}{sd(Y)\sqrt{(1-p)^2 Var(Z) + \frac{p}{n}Var(Z) + \frac{2p(1-p)r_{XZ}}{\sqrt{n}}Var(Z)}}$$

$$= \frac{(1-p)\mathrm{cov}(Y,Z)}{sd(Y)sd(Z)\sqrt{(1-p)^2 + \frac{p}{n} + \frac{2p(1-p)r_{XZ}}{\sqrt{n}}}}$$

$$+ \frac{p\,\mathrm{cov}(Y,X)}{sd(Y)\sqrt{n(1-p)^2 Var(X) + pVar(X) + \sqrt{n}\,2p(1-p)r_{XZ}Var(X)}}$$

$$= \frac{(1-p)r_{YZ}}{\sqrt{(1-p)^2 + \frac{p}{n} + \frac{2p(1-p)r_{XZ}}{\sqrt{n}}}} + \frac{pr_{XY}}{\sqrt{n(1-p)^2 + p + \sqrt{n}\,2p(1-p)r_{XZ}}}$$

Finally, as before the threshold of partial correlation$r_{XY|Z}$ -- $r^{\#}$, can be calculated to be

$$r^{\#} = \frac{t^{\#}}{\sqrt{t^{\#2} + res.df}},$$

where $t^{\#}$ is the critical value of t to invalidate inference, and res.df represents the residual degrees of freedom.

Thus to invalidate the observed inference we need to randomly rewire 100(1-p)% of ties in order to get

$$r^*_{XY|Z} = \frac{r^*_{XY} - r^*_{XZ}r_{YZ}}{\sqrt{1-r^{*\,2}_{XZ}}\sqrt{1-r^2_{YZ}}} = r^{\#}_{XY|Z}$$

As we can see, we can now represent the new partial correlation in terms of p (the percentage of ties retained), the original correlations in the observed data, and the average out-degree n. We can also write p as a function of other variables, but the formula is too complicated, so we do not provide it here.

For the anti-homophily rewiring we follow the same setup and let network exposure be X, the prior term be Z, the outcome be Y, and the average out-degree for each node be n. If we rewire (1-p)*100% of the ties of the nodes to the most dissimilar others, then assuming Z is centered at 0, which only affects the value of the new network exposure term X but not the correlation we are interested in, the new exposure term for actor i would become

$$X_i' = \frac{-n(1-p)Z_i + \sum_{j=1}^{np} Z_j}{n}.$$

And since $Var(X) = Var(\frac{\sum_{1}^{n} Z}{n}) = \frac{Var(Z)}{n}$, $Var(\frac{\sum_{1}^{np} Z}{n}) = \frac{p}{n}Var(Z)$. Also,

$Cov(Z, X') = -(1-p)Cov(Z, Z) + pCov(Z, X)$, and

$$Cov(X, Z) = sd(Z)sd(X)r_{XZ} = \frac{Var(Z)}{\sqrt{n}}r_{XZ}.$$

As a result, after rewiring the new correlation between network exposure and the prior become

$$r_{XZ}^{*} = \frac{\text{cov}(Z, X')}{sd(Z)sd(X')} = \frac{-(1-p)\,\text{cov}(Z, Z) + p\,\text{cov}(Z, X)}{sd(Z)sd(X')}$$

$$= \frac{-(1-p)\,\text{cov}(Z, Z) + p\,\text{cov}(Z, X)}{sd(Y)\sqrt{(1-p)^2 Var(Z) + \frac{p}{n}Var(Z) - 2p(1-p)Cov(X, Z)}}$$

$$= \frac{-(1-p)\,\text{cov}(Z, Z) + p\,\text{cov}(Z, X)}{sd(Y)\sqrt{(1-p)^2 Var(Z) + \frac{p}{n}Var(Z) - \frac{2p(1-p)r_{XZ}}{\sqrt{n}}Var(Z)}}$$

$$= \frac{-(1-p)\,\text{cov}(Z, Z)}{sd(Z)sd(Z)\sqrt{(1-p)^2 + \frac{p}{n} - \frac{2p(1-p)r_{XZ}}{\sqrt{n}}}}$$

$$+ \frac{p\,\text{cov}(Z, X)}{sd(Z)\sqrt{n(1-p)^2 Var(X) + pVar(X) - \sqrt{n}2p(1-p)r_{XZ}Var(X)}}$$

$$= \frac{-(1-p)}{\sqrt{(1-p)^2 + \frac{p}{n} - \frac{2p(1-p)r_{XZ}}{\sqrt{n}}}} + \frac{pr_{XZ}}{\sqrt{n(1-p)^2 + p - \sqrt{n}2p(1-p)r_{XZ}}}$$

Similarly, after rewiring, the new correlation between network exposure and the outcome become

$$r^{*}_{XY} = \frac{\mathrm{cov}(Y, X')}{sd(Y)sd(X')} = \frac{-(1-p)\mathrm{cov}(Y,Z) + p\,\mathrm{cov}(Y,X)}{sd(Y)sd(X')}$$

$$= \frac{-(1-p)\mathrm{cov}(Y,Z) + p\,\mathrm{cov}(Y,X)}{sd(Y)\sqrt{(1-p)^2 Var(Z) + \frac{p}{n}Var(Z) - 2p(1-p)Cov(X,Z)}}$$

$$= \frac{-(1-p)\mathrm{cov}(Y,Z) + p\,\mathrm{cov}(Y,X)}{sd(Y)\sqrt{(1-p)^2 Var(Z) + \frac{p}{n}Var(Z) - \frac{2p(1-p)r_{XZ}}{\sqrt{n}}Var(Z)}}$$

$$= \frac{-(1-p)\mathrm{cov}(Y,Z)}{sd(Y)sd(Z)\sqrt{(1-p)^2 + \frac{p}{n} - \frac{2p(1-p)r_{XZ}}{\sqrt{n}}}}$$

$$+ \frac{p\,\mathrm{cov}(Y,X)}{sd(Y)\sqrt{n(1-p)^2 Var(X) + pVar(X) - \sqrt{n}\,2p(1-p)r_{XZ}Var(X)}}$$

$$= \frac{-(1-p)r_{YZ}}{\sqrt{(1-p)^2 + \frac{p}{n} - \frac{2p(1-p)r_{XZ}}{\sqrt{n}}}} + \frac{pr_{XY}}{\sqrt{n(1-p)^2 + p - \sqrt{n}\,2p(1-p)r_{XZ}}}$$

Finally, as before the threshold for partial correlation$r_{XY|Z}$ -- $r^{\#}$, can be calculated to be

$$r^{\#} = \frac{t^{\#}}{\sqrt{t^{\#2} + res.df}},$$

where $t^{\#}$ is the critical value of t to invalidate inference, and res.df represents the residual degrees of freedom.

Thus to invalidate the observed inference we need to randomly rewire 100(1-p)% of ties in order to get

$$r^{*}_{XY|Z} = \frac{r^{*}_{XY} - r^{*}_{XZ}r_{YZ}}{\sqrt{1-r^{*2}_{XZ}}\sqrt{1-r^{2}_{YZ}}} = r^{\#}_{XY|Z}$$

As we can see we can now represent the new partial correlation in terms of p (the percentage of ties retained), the original correlations in the observed data, and the average out-degree n.

We can also write p as a function of other variables, but the formula is too complicated, so we do not provide it here.

B. Upper bound

Here we establish upper bound of social influence effects using estimation from homophily effects.

Estimated selection model

$$g(Z_{ijt}) = \gamma_0 + \gamma_1 In\deg ree_{jt-1} + \gamma_2 Similarity_{ij} + ...$$

Here, Z represents a network relationship. The term $Similarity_{ij}$ can be a composite measure of multiple attributes such as a cosine similarity:

$\cos(x_i, x_j) = \frac{\sum_k x_{ik} x_{jk}}{| x_i | \bullet | x_j |}$, where x_i is the vector of attributes for person i.

Then the relative magnitude of the standardized coefficient γ_2 represents the magnitude of relational balance, which is $\frac{\gamma_2}{\gamma_1 + \gamma_2 + ...}$. Let the influence model be

$$Y_{it} = \beta_0 + \beta_1 Y_{it-1} + \beta_2 \frac{\sum Z_{ijt-1} Y_{jt-1}}{\sum Z_{ijt-1}} + \beta_3 X_{it-1} + ... + e_{it}$$

Here X is a set of control variables, and the relative magnitude of standardized coefficient β_2 represents the magnitude of relational balance in the selection model, which is $\frac{\beta_2}{\beta_1 + \beta_2 + ...}$.

Then assuming that influence operates no faster than selection (which can be tested using empirical data), the upper-bound of $\frac{\beta_2}{\beta_1 + \beta_2 + ...}$ should be $\frac{\gamma_2}{\gamma_1 + \gamma_2 + ...}$.

If β_2 is over-estimated due to omitted variable bias, this upper bound can be useful in terms of determining the magnitude of bias.

It would be interesting to have multiple empirical data sets to test two things: (1) if the

homophily effect is the upper bound for social influence effects; (2) if homophily effects and social influence effects are indeed correlated.

C. Anti-homophily rewiring example

Here we give an anti-homophily rewiring example, N = 100. Density = 0.1, $r_{XY} > 0.2$. The prior term is a binary variable. Each point is a result of 1000 simulations. The analytical solutions in this case fit much better to the simulation results, compared with cases where the prior term is normally distributed.

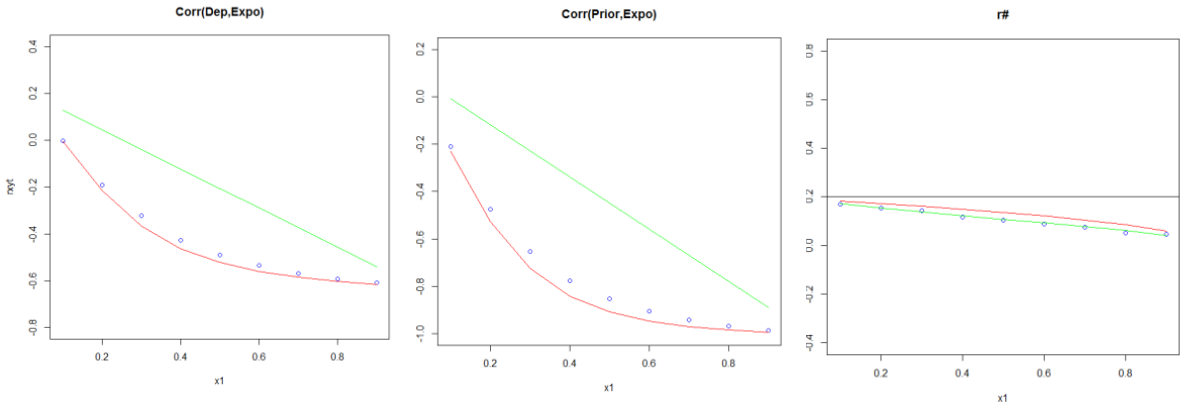

Figure S1: Anti-homophily rewiring example (better fit)

D. Homophily rewiring example

Here we provide some extra simulation examples for homophily rewiring, where they show fitting between analytical solutions and simulation results would be worse if we had a smaller or denser network. Example 1 in Figure S2 shows results for homophily rewiring when network is smaller, N = 50 instead of 100. Density = 0.1. Example 2 in Figure S3 shows results for homophily rewiring when network is denser, density = 0.2 instead of 0.1. N = 100.

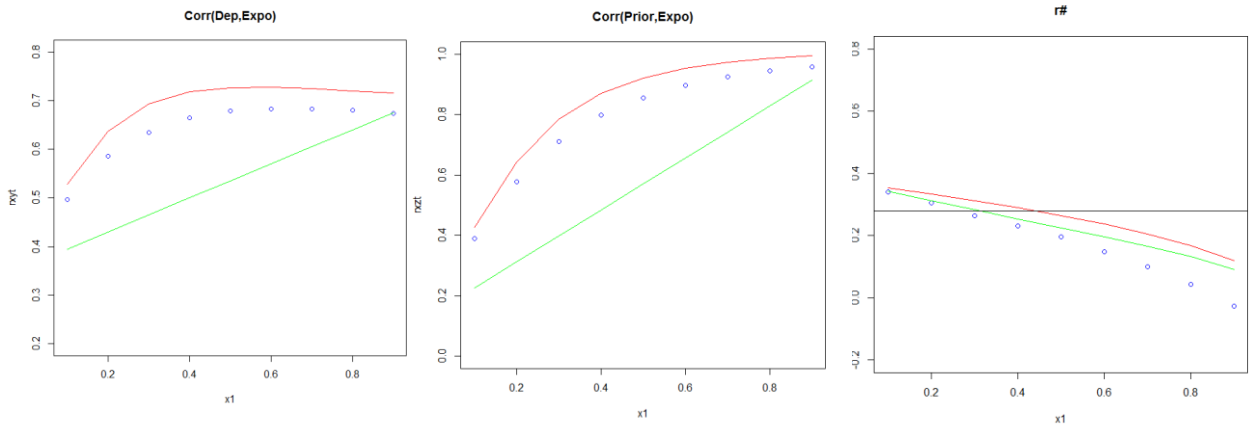

Figure S2: homophily rewiring example when network is smaller

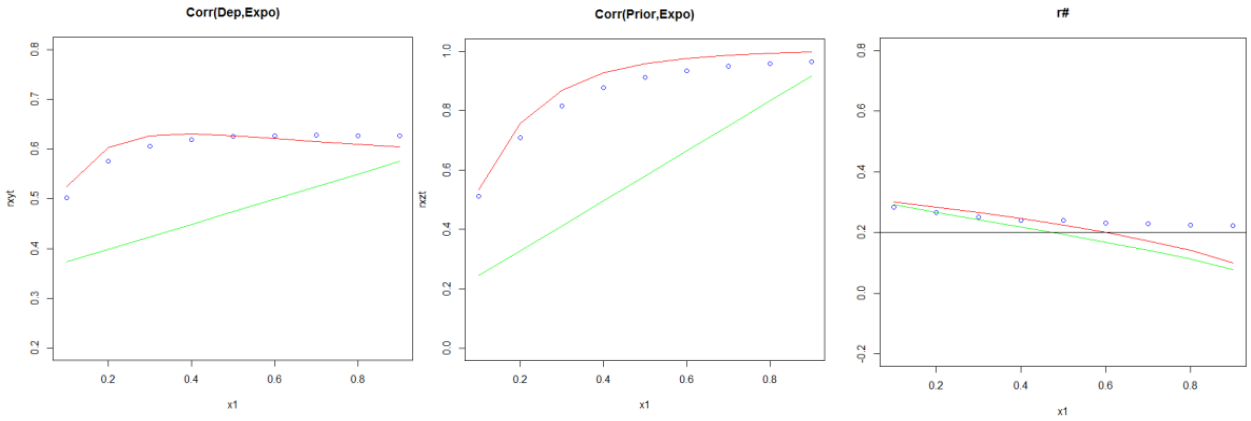

Figure S3: homophily rewiring example when network is dense