Determining the Effects of Social Network Evolution

Emily Muller

African Institute for Mathematical Sciences

August 31, 2017

AIMS SA Muizenberg

AIMS SA Muizenberg

Cosmology and Astrophysics, Mathematical and Physical Biosciences, Mathematical Finance and Mathematical Foundations and Scientific Computing.

2016/17 Structured Masters Students

2016/17 Structured Masters Students

A group is defined as "a finite set of actors who for conceptual, theoretical, or empirical reasons are treated as a finite set of individuals on which network measurements are made" (Wasserman and Faust, 1994).

2016/17 Structured MSc Social Network

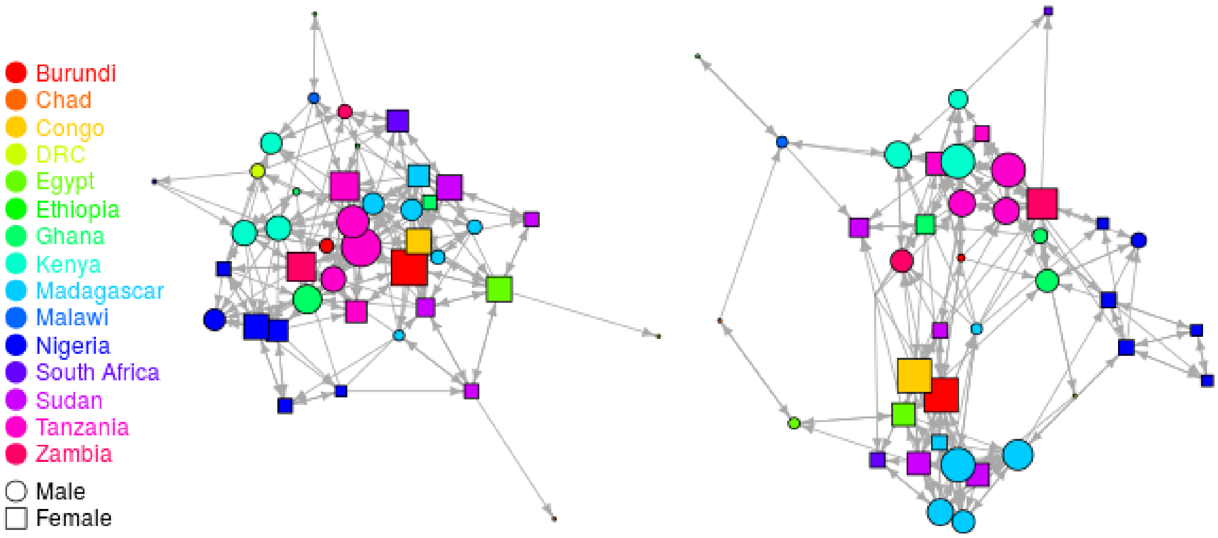


Figure: AIMS Network, December, $\boldsymbol{x}(t_0)$ (left) and April, $\boldsymbol{x}(t_1)$ (right). The size of each node is proportional to in-degree. The shape of the node represents the actors sex, and the colour of nodes represent country.

What are the **mechanisms** that determine social network evolution from t_0 to t_1?

History

History

Coleman, 1964

Continuous-time evolution although discrete time observations.

History

Coleman, 1964

Continuous-time evolution although discrete time observations.

Holland and Leindhart, 1981

Modelled evolution as continuous-time Markov process.

History

Coleman, 1964

Continuous-time evolution although discrete time observations.

Holland and Leindhart, 1981

Modelled evolution as continuous-time Markov process.

Wasserman, 1977

Continuous-time Markov Process with reciprocity as effect.

History

Coleman, 1964

Continuous-time evolution although discrete time observations.

Holland and Leindhart, 1981

Modelled evolution as continuous-time Markov process.

Wasserman, 1977

Continuous-time Markov Process with reciprocity as effect.

Snijders, 2001

Continuous-time Markov Process with multiple effects.

Continuous-Time Markov Process [Taylor and Karlin, 1998]

- $\mathcal{X} = \{\boldsymbol{x}(t) | t \in T\}$ is state space of all stochastic processes of order $2^{n(n-1)}$.
- $T = \{t \in \mathbb{R}^{+} | t_0 \leq t \leq t_1\}$.
- $P_{ij}(t, s) = P[\boldsymbol{x}(t + s) = j | \boldsymbol{x}(t) = i]$.
- $P_{ij}(t, s) = \prod_{k,l} P\{\boldsymbol{x}_{kl}(t + s) = j_{kl} | \boldsymbol{x}(t) = i\} + o(s)$.
- $P_{ij}(t, s) = P[\boldsymbol{x}(s) = j | \boldsymbol{x}(0) = i] = P_{ij}(s)$.
- Regularity.
- Infinitesimal Generator, q_{ij}, *rate of change of transition*.
- Markov process: initialise $\boldsymbol{x}(0) = i$. Soujourn in state i for a time exponentially distributed with parameter q_{ii}. Transition to state j with probability $p_{ij} = \frac{q_{ij}}{q_{ii}}$ and repeat.

Stochastic Actor-Oriented Model [Snijders, 2001]

- Continuous-time.
- Markovian network.
- Discrete choice model, $\boldsymbol{x}(i \rightsquigarrow j)$.

Choices/Mechanisms/Effects

Assume the choice to make or break a tie with $n-1$ other actors in the network is individuals own choice and dependent on network and covariate effects.

TABLE 2
Selection of Possible Effects for Modeling Network Evolution

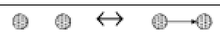
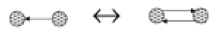
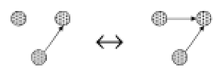
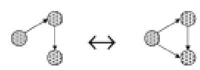
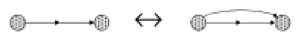
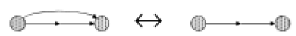
Effect	Network Statistic	Effective Transitions in Network[a]	Verbal Description
1. Outdegree	$\sum_j \mathbf{x}_{ij}$		Overall tendency to have ties
2. Reciprocity	$\sum_j \mathbf{x}_{ij}\mathbf{x}_{ji}$		Tendency to have reciprocated ties
3. Preferential attachment	$\sum_j \mathbf{x}_{ij}\sqrt{\sum_h \mathbf{x}_{hj}}$		Tendency to attach to popular others (with decreasing marginal sensitivity to alter's popularity)
4. Transitive triplets	$\sum_j \mathbf{x}_{ij} \sum_h \mathbf{x}_{ih}\mathbf{x}_{hj}$		Tendency toward triadic closure of the neighborhood (linear effect of the number of indirect ties)
5. Transitive ties	$\sum_j \mathbf{x}_{ij} \max_h(\mathbf{x}_{ih}\mathbf{x}_{hj})$	 (*number* of intermediaries is irrelevant)	Tendency toward triadic closure of the neighborhood (binary effect of indirect ties)
6. Actors at distance 2	$\sum_j (1-\mathbf{x}_{ij}) \max_h(\mathbf{x}_{ih}\mathbf{x}_{hj})$	 (*number* of intermediaries is irrelevant)	Tendency to keep others at social distance 2 (negative measure of triadic closure)

Figure: *Dynamic Networks and Behaviour: Separating Selection From Influence*, Steglich et al., 2010.

Discrete Choice Model

Rate Function: rate at which actor i chooses to select or deselect friendship. Assume constant: λ.

Objective Function: Perceived utility, of actor i, for chosen network configuration, $\boldsymbol{x}(i \rightsquigarrow j)$,

$$f(i, \boldsymbol{x}(i \rightsquigarrow j)) = \sum_{s=1}^{L} \beta_s \rho_s(i, \boldsymbol{x}(i \rightsquigarrow j)).$$

Discrete Choice Model

Rate Function: rate at which actor i chooses to select or deselect friendship. Assume constant: λ.
Objective Function: Perceived utility, of actor i, for chosen network configuration, $\boldsymbol{x}(i \rightsquigarrow j)$,

$$f(i, \boldsymbol{x}(i \rightsquigarrow j)) = \sum_{s=1}^{L} \beta_s \rho_s(i, \boldsymbol{x}(i \rightsquigarrow j)).$$

Discrete Choice Model [Maddala, 1983]:

$$max_{j \in V(\boldsymbol{x})}(f(i, \boldsymbol{x}(i \rightsquigarrow j)) + U_j),$$

where U_j (assumed i.i.d Gumbel). Probability multinomial logistic regression given by

$$p_{ij}(\boldsymbol{x}(i \rightsquigarrow j)) = \frac{exp(f(i, \boldsymbol{x}(i \rightsquigarrow j)) - f(i, \boldsymbol{x}))}{\sum_{h=1, h \neq i}^{n} exp(f(i, \boldsymbol{x}(i \rightsquigarrow h)) - f(i, \boldsymbol{x}))}.$$

Markov Process

A Markov process is completely defined by the space of all possible states $\mathcal{X}$, the initial state, $\boldsymbol{x}(t_0)$, and the transition rate matrix Q,

$$q_{ij} = \lambda p_{ij}.$$

Model is dependent on unknown parameters $\hat{\boldsymbol{\theta}} = (\hat{\lambda}, \hat{\boldsymbol{\beta}})$.

1. Initialise: $t = 0, \boldsymbol{x}(t_0)$, $\boldsymbol{\rho}$ and $\hat{\boldsymbol{\theta}}$.
2. Sample i from uniform distribution.
3. Given actor i, sample j with probability $p_{ij}(\boldsymbol{x}(i \rightsquigarrow j))$.
4. Let $t = t + \Delta t$ for Δt sampled exponential random variable with parameter $n\hat{\lambda}$.
5. Change network $\boldsymbol{x}(t)(i \rightsquigarrow j)$.
6. Repeat step (b) until $t = T_1$.

Denote the final output $\boldsymbol{x}(T_1)$. $\boldsymbol{x}$ is therefore dependent on T_1, $\boldsymbol{x}(t_0)$, $\boldsymbol{\rho}$ and $\hat{\boldsymbol{\theta}}$.

Parameter Estimation

Method of Moments [Snijders, 2001]:

$$E[Z(\boldsymbol{x}(T_1, \hat{\boldsymbol{\theta}}))|\boldsymbol{x}(t_0), \boldsymbol{\rho}] = z^{obs},$$

for $\hat{\boldsymbol{\theta}} = (\hat{\lambda}, \hat{\boldsymbol{\beta}})$. Chose $Z = (C(t), \boldsymbol{P}(t))$ where

$$C = ||\boldsymbol{x}(t) - \boldsymbol{x}(t_0)|| = \sum_{1 \le i,j \le n} |X_{ij}^{t} - X_{ij}^{t_0}|,$$

$$P_s = \sum_{i=1}^{n} \rho_s(\boldsymbol{x}(t), i), \text{ for } s \in [1, L].$$

Parameter Estimation

Method of Moments [Snijders, 2001]:

$$E[Z(\boldsymbol{x}(T_1, \hat{\boldsymbol{\theta}}))|\boldsymbol{x}(t_0), \boldsymbol{\rho}] = z^{obs},$$

for $\hat{\boldsymbol{\theta}} = (\hat{\lambda}, \hat{\boldsymbol{\beta}})$. Chose $Z = (C(t), \boldsymbol{P}(t))$ where

$$C = ||\boldsymbol{x}(t) - \boldsymbol{x}(t_0)|| = \sum_{1 \le i,j \le n} |X_{ij}^{t} - X_{ij}^{t_0}|,$$

$$P_s = \sum_{i=1}^{n} \rho_s(\boldsymbol{x}(t), i), \text{ for } s \in [1, L].$$

The moment equations are

$$E[C(\boldsymbol{x}(T_1, \hat{\boldsymbol{\theta}}))|\boldsymbol{x}(t_0), \boldsymbol{\rho}] = c^{obs}, \quad (1)$$

$$E[\boldsymbol{P}(\boldsymbol{x}(T_1, \hat{\boldsymbol{\theta}}))|\boldsymbol{x}(t_0), \boldsymbol{\rho}] = \boldsymbol{p}^{obs}. \quad (2)$$

Conditional Moment Estimation

$$E[C(\boldsymbol{x}(T_1, \hat{\boldsymbol{\theta}}))|\boldsymbol{x}(t_0), \boldsymbol{\rho}] = c^{obs} = ||\boldsymbol{x}(t_1) - \boldsymbol{x}(t_0)||.$$

The expected number of changes in the simulated network must be equal to the number of changes in the observed network (from initial network).
Impose the following: $T_1 = \min\{t|C(t) \geq c^{obs}\}$.
Moment equation is

$$E[\boldsymbol{P}(\boldsymbol{x}(T_1, \hat{\boldsymbol{\beta}}))|\boldsymbol{x}(t_0), \boldsymbol{\rho}, C] = \boldsymbol{p}^{obs}.$$

Stochastic Approximation

Snidjers uses an updated version of the Robbins-Monro [1951] method to iteratively update the parameters

$$\hat{\boldsymbol{\beta}}_{N+1} = \hat{\boldsymbol{\beta}}_N - a_N D_0^{-1}(\boldsymbol{P}_N - \boldsymbol{p}^{obs})$$

where N is the step in the CTMP, a_N is a series that slowly converges to 0 with rate N^{-c} $(0.5 < c < 1)$ and D_0 is the diagonal matrix with entries: $D_{\hat{\boldsymbol{\beta}}_t} = \dfrac{\partial E[\boldsymbol{P}]}{\partial \hat{\boldsymbol{\beta}}_t}$ on the diagonal.

AIMS African Institute for Mathematical Sciences SOUTH AFRICA

Stochastic Approximation

Snidjers uses an updated version of the Robbins-Monro [1951] method to iteratively update the parameters

$$\hat{\boldsymbol{\beta}}_{N+1} = \hat{\boldsymbol{\beta}}_N - a_N D_0^{-1}(\boldsymbol{P}_N - \boldsymbol{p}^{obs})$$

where N is the step in the CTMP, a_N is a series that slowly converges to 0 with rate N^{-c} $(0.5 < c < 1)$ and D_0 is the diagonal matrix with entries: $D_{\hat{\boldsymbol{\beta}}_t} = \dfrac{\partial E[\boldsymbol{P}]}{\partial \hat{\boldsymbol{\beta}}_t}$ on the diagonal.

Optimal convergence:

- Polyak [1996]: when D_0 has positive real eigenvalues and $\hat{\boldsymbol{\beta}}_N$ generated by average of consecutive values.

To have good convergence for relatively low N:

- Pflug [1990] showed $\boldsymbol{P_N} - \boldsymbol{p}^{obs}$ negative.
- If $(\boldsymbol{P_N} - \boldsymbol{p}^{obs})'(\boldsymbol{P_{N-1}} - \boldsymbol{p}^{obs})$ positive then drifting towards limit point and a_N remains constant.

MCMC

- Phase I: approximate D_0 using common random numbers.
- Phase II: subphases κ with constant a_N. Bounded by positive successive products and steps, $(n_{2\kappa}^{-}, n_{2\kappa}^{+})$, so that $N^{3/4}a_N$ tends to positive finite limit. At the end of each subphase the average estimate is used as input for next subphase. $\hat{\boldsymbol{\beta}}_N$ over last subphase is used as final output $\hat{\boldsymbol{\beta}}$.
- Phase III: Given $\hat{\boldsymbol{\beta}}$ estimate $cov(\hat{\boldsymbol{\beta}}) \approx D_{\hat{\boldsymbol{\beta}}}^{-1} \Sigma_{\hat{\boldsymbol{\beta}}} D_{\hat{\boldsymbol{\beta}}}'^{-1}$.

  1000 networks generated: $\boldsymbol{x}(T_1, \hat{\boldsymbol{\beta}})$.

2016/17 Structured MSc Social Network

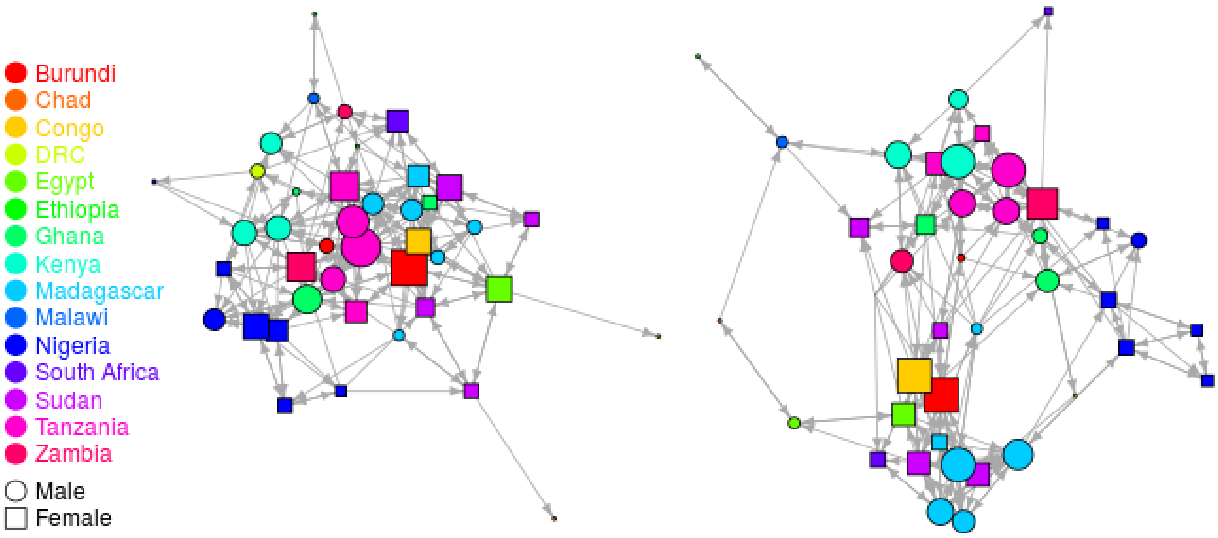


Figure: AIMS Network, December, $\boldsymbol{x}(t_0)$ (left) and April, $\boldsymbol{x}(t_1)$ (right). The size of each node is proportional to in-degree. The shape of the node represents the actors sex, and the colour of nodes represent country.

2016/17 Structured MSc Social Network

Table: Network Topologies for AIMS Network

	December	April
Number of Nodes	41	41
Number of Edges	212	203
Density	0.13	0.12
Average Degree	5.17	4.95
Reciprocity	0.58	0.58
Transitiviy	0.30	0.39
Distance	2.58	2.96

The network has changed by a total of $c = 169$ ties. Jaccard coefficient

$$\frac{N_{11}}{N_{11} + N_{01} + N_{10}} = 0.4.$$

Model Implementation

Table: Parameter Estimation of Friendship Evolution for AIMS Network.

	Network Effects	Model I				Model II				Model III			
		Estimate	S.E	p-value	Conv. t-ratio	Estimate	S.E	p-value	Conv. t-ratio	Estimate	S.E	p-value	Conv. t-ratio
0	Rate parameter	7.61	(0.85)			7.70	(0.86)			7.69	(0.88)		
1 .	eval outdegree (density)	-1.40	(0.24)	3.66e-09	-0.04	-1.33	(0.17)	1.71e-15	0.02	-1.26	(0.18)	5.46e-12	-0.04
2 .	eval reciprocity	1.38	(0.22)	7.12e-10	-0.01	1.32	(0.19)	1.18e-11	0.05	1.30	(0.21)	2.99e-10	-0.02
3 .	eval transitive triplets	0.21	(0.08)	0.01	-0.02	0.18	(0.05)	5.37e-04	0.03	0.17	(0.05)	6.99e-04	-0.07
4 .	eval 3-cycles	-0.03	(0.13)	0.79	0.02								
5 .	eval transitive ties	-0.04	(0.23)	0.88	-0.01								
6 .	eval balance	0.04	(0.02)	0.03	0.03	0.04	(0.02)	0.03	-0.01	0.05	(0.02)	0.01	0.06
7 .	eval number of actors at distance 2	-0.27	(0.07)	1.10e-3	-0.03	-0.27	(0.07)	1.10e-04	-0.03	-0.29	(0.07)	4.68e-05	-0.03
10 .	eval same country	0.58	(0.24)	0.01	-0.01	0.57	(0.23)	0.01	0.01	0.54	(0.24)	0.03	0.03
11.	eval sex alter	0.20	(0.16)	0.18	0.07								
12.	eval sex ego	0.04	(0.16)	0.81	0.06								
13.	eval same sex	0.05	(0.17)	0.76	-0.03								

Network Topology

Table: Global network metrics, $\mu_q(\boldsymbol{x})$.

μ_q	$\boldsymbol{x}$
μ_1: Density	$m/n(n-1)$
μ_2: Reciprocity	$\sum_{1\leq j,k\leq n} X_{ij}X_{ji}/m$
μ_3: Global clustering	$\frac{\#\text{transitive triplets}}{\#\text{ of connected triplets of vertices}}$
μ_4': Harmonic mean distance	$n/\sum_{i=1}^{n}\gamma_4(\boldsymbol{x},i)$

Table: Local network metrics, $\gamma_r(\boldsymbol{x},i)$.

γ_r	$\boldsymbol{x}$
γ_1: Out-degree dist.	
γ_2: In-degree dist.	
γ_3: Local clustering	$\sum_{1\leq l,m\leq k_i} X_{lm}/k_i(k_i-1)$
γ_4': Harmonic closeness	$1/(n-1)\sum_{j=1,j\neq i}^{n}\frac{1}{d_{ij}}$

Method of Moments

$P_s = \sum_{i=1}^{n} \rho_s(\boldsymbol{x}(T_1), i), \text{ for } s \in [1, L].$

Table: P_s and p_s statistics for network effects, ρ_s, for Model III

Effect	ρ_s	Target p_s	Mean Estimate P_s
Out-degree	ρ_1	203	202
Reciprocity	ρ_3	118	118
Transitive triplets	ρ_4	360	355
Number Distance 2	ρ_6	429	429
Balance	ρ_8	399	405
Same country	$\rho_{15}^{country}$	65	65

Network Difference: μ_q

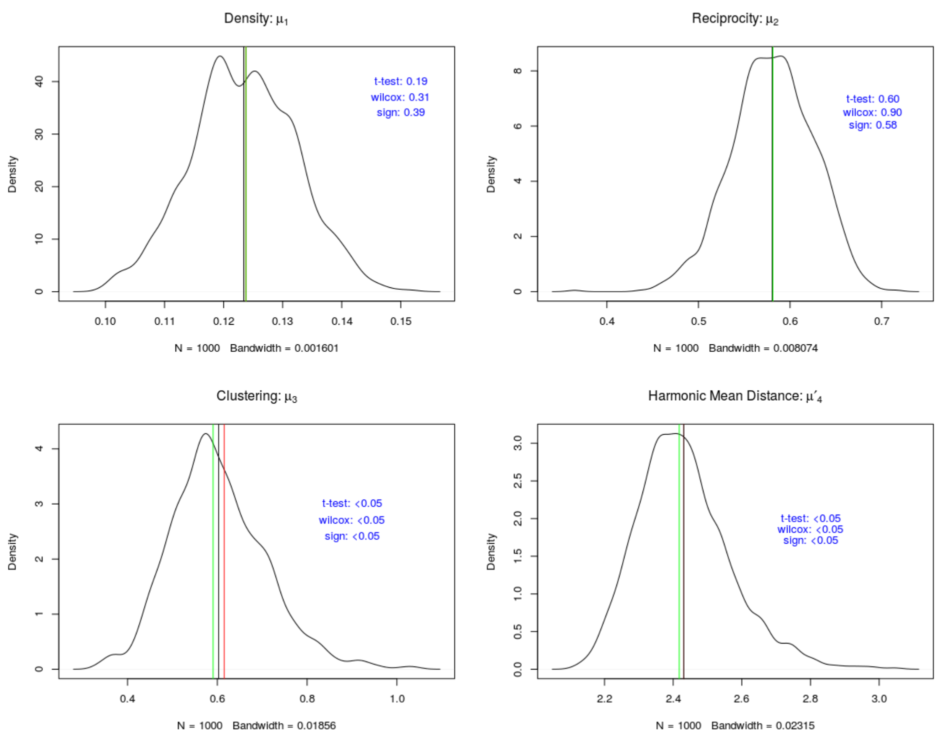


Figure: Density plots for difference metrics, μ_q. Red line is μ^{obs}, black line is $\bar{\mu}$ and green line is μ^{median}.

Network Difference: γ_q

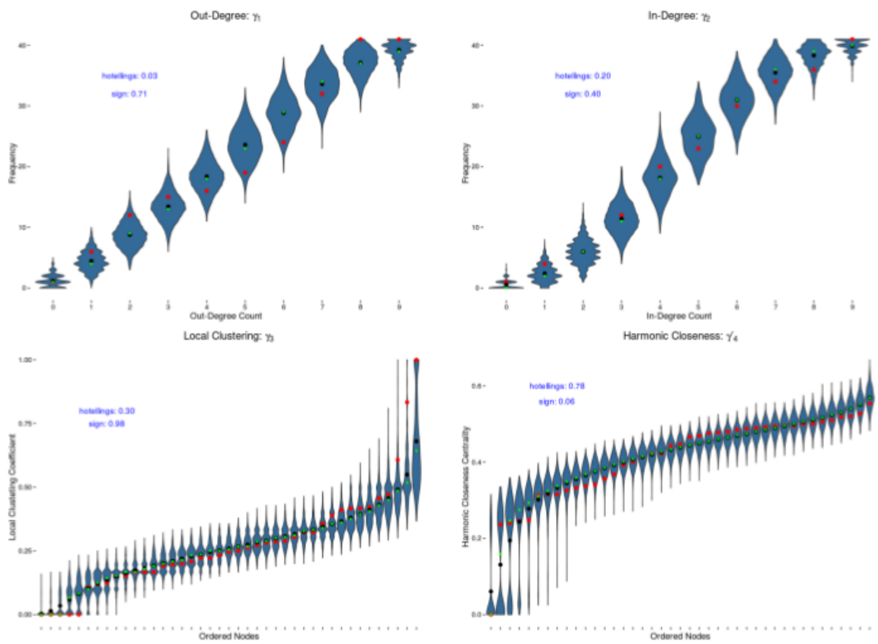


Figure 5.4: Violin Plots for difference metrics, γ_r, red dots are γ^{obs} and green dots are γ^{median}. The blue shaded region (violin) represents the sample data for each variable. In-degree and out-degree show cumulative degree distribution. The width of the violin is proportional to the density and the length is proportional to the range. Symmetry of the data is indicated by symmetrical violins on the horizontal and vertical axis.

Research Questions

- How do well-known principles of network formation, namely reciprocity, popularity, and triadic closure, vary in importance throughout the network formation period as the structure itself evolves? (Schaefer, Light, Fabes, Hanish, & Martin, 2010)
- How does peer influence on smoking cessation differ in magnitude from peer influence on smoking initiation? (Haas & Schaefer, 2014)
- What drives collaboration among collective actors involved in climate mitigation policy? (Ingold & Fischer, 2014)
- Why are some more peer than others? evidence from a longitudinal study of social networks and individual academic performance. (Lomi, Snijders, Steglich & Torló, 2011)

Model Limitations

- Markovian, one tie change...
- Constant effects for more than two observations.
- Closed group study.
- Expensive data collection.
- Accuracy and reliability - inference.

$\Longrightarrow$ Online social networks.

Thank you.