STATISTICAL HORIZONS

Social Networks: Statistical Approaches

John Skvoretz, Ph.D.

Upcoming Seminar:
November 15-16, 2019, Philadelphia, Pennsylvania

Session 1.1:
Hypotheses about Reciprocity, Multiplexity, Exchange, Transitivity, Density, Degree, Centralization, Clustering, and Other Graph-level Indices

John Skvoretz
University of South Florida
jskvoretz@usf.edu

Session questions

- How do patterns of ties in dyads and triads differ from chance?
- How can we statistically test for forces structuring dyads and triads in actual networks?
- How can we tell if a network property like density or betweenness centralization differs significantly from chance expectations?
- How can we tell if two networks differ significantly in some property, like their clustering coefficients?

Topics covered

- Structural forces in dyads
  - Reciprocity – one type of tie
  - Multiplexity – two or more types of ties
  - Exchange – two or more types of ties
  - Reciprocity, multiplexity, exchange within and between subgroups and groups

Topics covered

- Structural forces in triads
  - Transitivity/closure/similarity
  - Intransitivity
  - Openness
  - Dissimilarity

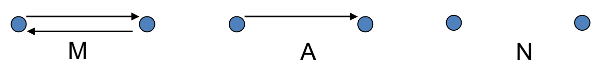
Topics covered

- Common graph metrics
- Simple statistical tests
- Tests in the `sna` package
- Less common graph metrics

Structural forces in dyads – reciprocity/

Reciprocity

- Reciprocity (also called mutuality)
  - “a tendency to reciprocate choices more frequently than is indicated by chance coincidence of choice” Katz and Powell (1955: 659)
  - Revealed by the dyad census, the number of M, A, and N dyads observed compared to the number expected by chance



Reciprocity

- How many mutual dyads could we expect due to “chance coincidence of choice”?
- Assume each person in a group of size N makes d choices from his/her N-1 possible partners
- The probability M occurs in any one dyad is the product of the probability i chooses j and the probability j chooses i

$$\Pr(i \to j)\Pr(j \to i) = \frac{d}{N-1}\frac{d}{N-1} = \frac{d^2}{(N-1)^2}$$

Reciprocity

- The number of mutuals expected by chance is this probability multiplied by the number of dyads:

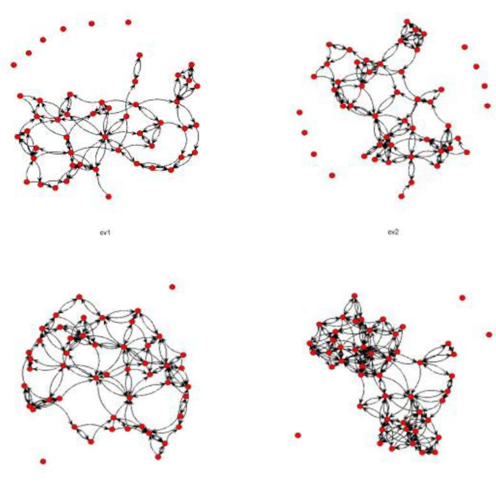
$$M_{exp} = \frac{N(N-1)}{2}\frac{d^2}{(N-1)^2} = \frac{Nd^2}{2(N-1)}$$

- The other probabilities and expected numbers are

$$\Pr(A) = \frac{2d(N-1-d)}{(N-1)^2} \qquad A_{exp} = \frac{Nd(N-1-d)}{(N-1)}$$

$$\Pr(N) = \frac{(N-1-d)^2}{(N-1)^2} \qquad N_{exp} = \frac{N(N-1-d)^2}{2(N-1)}$$

Reciprocity



Reciprocity

	Observed			
	cv1	cv2	cv3	cv4
M	22	38	47	66
A	60	41	56	71
N	1143	1146	1122	1088
Mean Degree	2.08	2.34	3.00	4.06

	Expected			
	2	3	4	5
M	2.04	4.59	8.16	12.76
A	95.92	140.82	183.67	224.49
N	1127.04	1079.59	1033.16	987.76
Mean Degree	2.00	3.00	4.00	5.00

An index for reciprocity

- General form of Katz and Powell's tau index

$$\tau = \frac{M_{obs} - M_{\exp}}{M_{\max} - M_{\exp}}$$

- When all make d choices, then

$$M_{\max} = \frac{Nd}{2} \quad \text{and so}$$

$$\tau = \frac{M_{obs} - \dfrac{Nd^2}{2(N-1)}}{\dfrac{Nd}{2} - \dfrac{Nd^2}{2(N-1)}} = \frac{2(N-1)M_{obs} - Nd^2}{Nd(N-1-d)}$$

An index for reciprocity

- If M_{obs}=M_{exp} then tau=0 and there is no force towards reciprocation
- If M_{obs}=M_{max} then tau=1 and the force towards reciprocation is at its maximum
- Reciprocity takes choices that would by chance be sent to those who did not choose i and "redirects" them to those who choose i:

$$\Pr(\text{iMj}) = \frac{d}{N-1}\left[\frac{d}{N-1} + \tau\left(1 - \frac{d}{N-1}\right)\right]$$

An index for reciprocity

- Other "chance" models can be used to define M_{exp} and M_{max}
- Conditional on the total number of ties – all networks with exactly the same total number of ties are equally likely

$$\tau \mid x_{++}^{A} = \frac{M_{obs} - \frac{\left(x_{++}^{A}\right)^2 - x_{++}^{A}}{2(N(N-1)-1)}}{\left\lfloor x_{++}^{A}/2 \right\rfloor - \frac{\left(x_{++}^{A}\right)^2 - x_{++}^{A}}{2(N(N-1)-1)}}$$

An index for reciprocity

- Conditional on the outdegree distribution (the “free choice” case) – all networks with exactly the same outdegree distribution are equally likely

$$\tau \mid \langle x_{i+}^{A} \rangle = \frac{M_{obs} - \dfrac{\left(x_{++}^{A}\right)^{2} - \sum_{i=1}^{N}\left(x_{i+}^{A}\right)^{2}}{2(N-1)^{2}}}{\left[\dfrac{x_{++}^{A} - \max\limits_{0 \le t \le N}\left[\sum_{i=1}^{t} x_{i+}^{A} - t(t-1) - \sum_{i=t+1}^{N} \min(t, x_{i+}^{A})\right]}{2}\right] - \dfrac{\left(x_{++}^{A}\right)^{2} - \sum_{i=1}^{N}\left(x_{i+}^{A}\right)^{2}}{2(N-1)^{2}}}$$

Some observations on the index

- No chance model conditional on indegree distribution is proposed
  - Because “the numbers of choices received are a side-effect, as it were, of the strength of reciprocal tendency” (Katz and Powell 1955: 662)
- The index has “heuristic” value
  - Does not itself provide a statistical test for whether there is significantly more reciprocity than expected by chance

Statistical test for reciprocity

- Statistical test compares observed to chance
  - Calculate the expected index score under a specific chance distribution
    - Easy – it is always 0!
  - Calculate the variance in the expected score under the chance distribution
    - Tricky because dyads (such as, ij and jk) are not independent

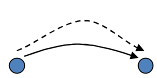
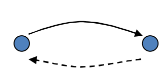
Statistical test for reciprocity

- Calculate a z-score to express the difference between the observed index score and the expected score in terms of standard deviations

$$\text{z-score} = \frac{\tau_{obs} - \tau_{\exp}}{\sqrt{Var(\tau_{\exp})}} = \frac{\tau_{obs}}{\sqrt{Var(\tau_{\exp})}} = \frac{\tau_{obs}}{\sqrt{\frac{Var(M_{\exp})}{(M_{max} - M_{\exp})^2}}}$$

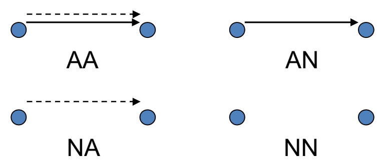
Structural forces in dyads – multiplexity and exchange/

Multiplexity and exchange

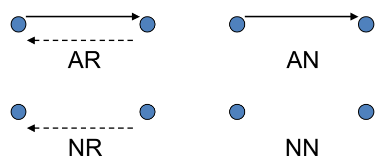
- Multiplexity – a tendency for a tie of type A from i to j to co-occur with a tie of type B from i to j at greater than chance levels
  - friendship and seeks advice from 
- Exchange – a tendency for a tie of type A from i to j to co-occur with a tie of type B from j to i at greater than chance levels
  - gives advice to and gets respect from 

Multiplexity and exchange

- To detect departures from chance, use census of pairs
  - For multiplexity



  - For exchange



Multiplexity and exchange

- Under a force towards multiplexity
  - Configurations with pairs having both an A and B tie in the same direction will occur at greater than chance levels
- Under a force towards exchange
  - Configurations with pairs have an A tie in one direction and a B tie in the other will occur at greater than chance levels

Multiplexity and exchange

- Define indices similar to K&P's tau

$$\upsilon = \frac{Multi_{obs} - Multi_{\exp}}{Multi_{\max} - Multi_{\exp}} \qquad \varepsilon = \frac{Exch_{obs} - Exch_{\exp}}{Exch_{\max} - Exch_{\exp}}$$

- Define chance models to calculate expected and maximum
  - Conditional on total numbers of A and B ties
  - Conditional on joint outdegree distribution of A and B ties

Multiplexity and exchange

- Chance model 1 – condition on total A and B ties
  - Maximum for both multiplexity and exchange is the smaller of the number of A and B ties

$$Mult_{max} = \min\left[x^A_{++}, x^B_{++}\right] = Exch_{\max}$$

  - Expected number for both is the total number of pairs multiplied by the probability any one is multiplex (or in exchange)

$$Mult_{\exp} = N(N-1)\left(\frac{x^A_{++}}{N(N-1)}\right)\left(\frac{x^B_{++}}{N(N-1)}\right) = Exch_{\exp}$$

Multiplexity and exchange

- Chance model 2 – condition on the A and B outdegree distributions (multiplexity and exchange differ)
  - Maximum for multiplexity is the sum over i of the smaller of i's A and B ties

$$Mult_{max} = \sum_i \min\left[x_{i+}^A, x_{i+}^B\right]$$

  - Expected for multiplexity is the sum over i of the expected number of i's ties that are multiplex

$$Mult_{\exp} = \sum_i \sum_{j \neq i} \frac{x_{i+}^A x_{i+}^B}{(N-1)^2} = \sum_i \frac{x_{i+}^A x_{i+}^B}{(N-1)}$$

Multiplexity and exchange

  - Expected for exchange is the sum over all pairs i≠j of the chance i sends a tie to j and j sends a tie to i:

$$Exch_{\exp} = \sum_{i \neq j} \frac{x_{i+}^A x_{j+}^B}{(N-1)}$$

  - But the maximum for exchange is much more complicated and depends greatly on the exact degree sequences:

$$Exch_{\max} = \sum_{i=1}^{N} \min\left[x_{i+}^A, \#\{x_{j+,k}^{B[i-1]} \mid x_{j+,k}^{B[i-1]} > 0 \wedge j \neq i\}\right]$$

Statistical tests

- Statistical tests for multiplexity and exchange compare observed to chance
  - Note the expected values of υ and ε under any chance distribution are zero
  - Also variances are functions of variances in the number of multiplex or exchange pairs
  - Form z-scores

$$\text{z-score} = \frac{\upsilon_{obs}}{Var(\upsilon_{\exp})} = \frac{\upsilon_{obs}}{\sqrt{\dfrac{Var(Mult_{\exp})}{(Mult_{max} - Mult_{exp})^2}}} \qquad \text{z-score} = \frac{\varepsilon_{obs}}{Var(\varepsilon_{\exp})} = \frac{\varepsilon_{obs}}{\sqrt{\dfrac{Var(Exch_{\exp})}{(Exch_{max} - Exch_{exp})^2}}}$$

Group comparisons

- Comparing subgroups and two separate groups (Agneessens and Skvoretz 2012)
- For a population divided into two subgroups,
  - Compare reciprocity, multiplexity, and exchange scores within each group to each other and to the scores between groups
- For two separate groups,
  - Compare reciprocity, multiplexity, and exchange indices in one group to those in the other group