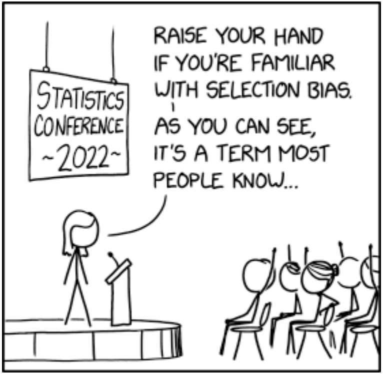
Network Statistics

Lorien Jasny and Örjan Bodin



Hypothesis Testing

Hypothesis Testing

- Can't use many existing statistical methods

Hypothesis Testing

- Can't use many existing statistical methods
  - Permutation tests (eg QAP) and baseline models
  - Network Autocorrelation Models
  - Exponential Random Graph models (ERGM)

Hypothesis Testing

- Can't use many existing statistical methods
    - Permutation tests (eg QAP) and baseline models
    - Network Autocorrelation Models
    - Exponential Random Graph models (ERGM)
- Model choice depends (mostly) on your selection of dependent/independent variables

Relating Node level indices to covariates

- Node Level Indices: centrality measures, brokerage, constraint

Relating Node level indices to covariates

- Node Level Indices: centrality measures, brokerage, constraint
- Node Covariates: measures of power, career advancement, gender – really anything you want to study that varies at the node level

Emergent Multi-Organizational Networks (EMON) Dataset

- 7 case studies of EMONs in the context of search and rescue activities from Drabek et. al. (1981)

Emergent Multi-Organizational Networks (EMON) Dataset

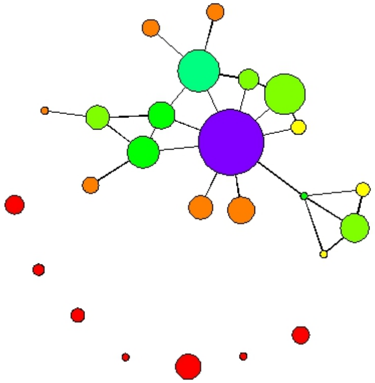
- 7 case studies of EMONs in the context of search and rescue activities from Drabek et. al. (1981)
- Ties between organizations are self-reported levels of communication coded from 1 to 4 with 1 as most frequent

Emergent Multi-Organizational Networks (EMON) Dataset

Attribute Data

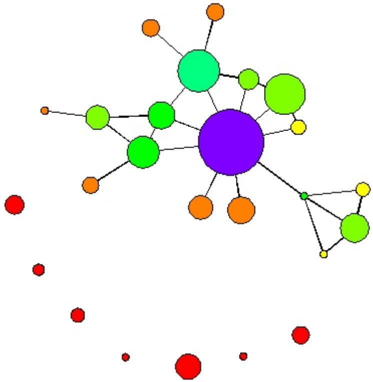
- Command Rank Score (CRS): mean rank (reversed) for prominence in the command structure
- Decision Rank Score (DRS): mean rank (reversed) for prominence in decision making process
- Paid Staff: number of paid employees
- Volunteer Staff: number of volunteer staff
- Sponsorship: organization type (City, County, State, Federal, or Private)

Correlation between DRS and Degree?



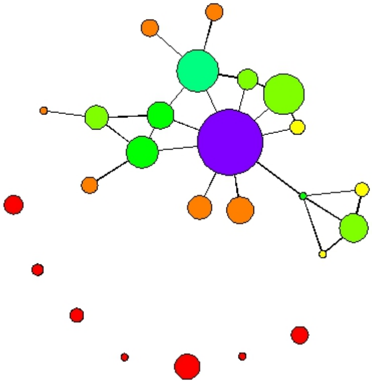
Correlation between DRS and Degree?

- Subsample of Mutually Reported "Continuous Communication" in Texas EMON



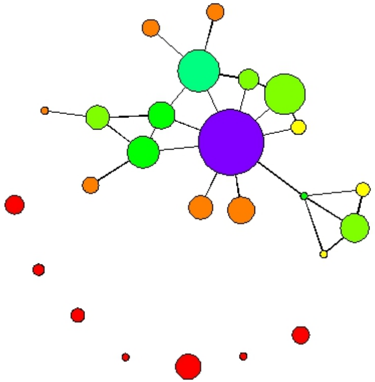
Correlation between DRS and Degree?

- Subsample of Mutually Reported "Continuous Communication" in Texas EMON
- Degree is shown in color (darker is bigger)



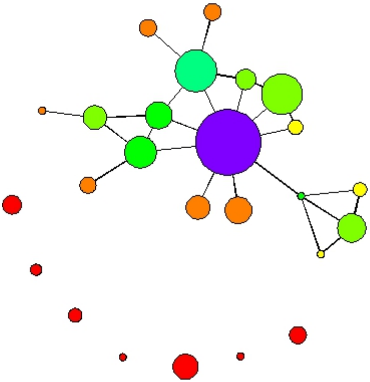
Correlation between DRS and Degree?

- Subsample of Mutually Reported "Continuous Communication" in Texas EMON
- Degree is shown in color (darker is bigger)
- DRS in size

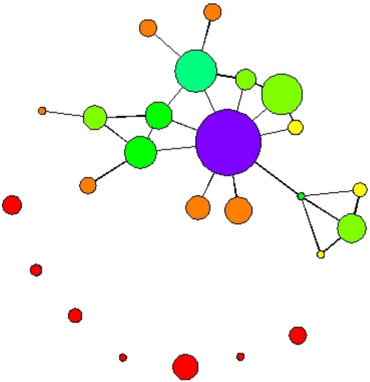


Correlation between DRS and Degree?

- Subsample of Mutually Reported "Continuous Communication" in Texas EMON
- Degree is shown in color (darker is bigger)
- DRS in size
- Empirical corelation $\rho = 0.86$

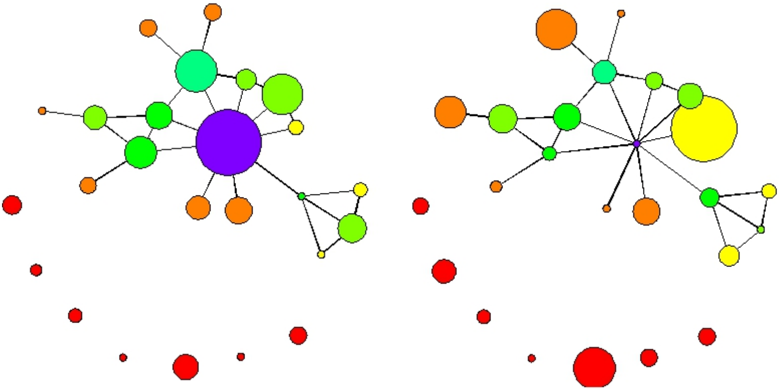


Correlation between DRS and Degree?



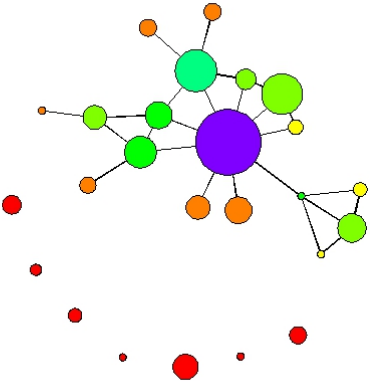
$\rho = 0.86$

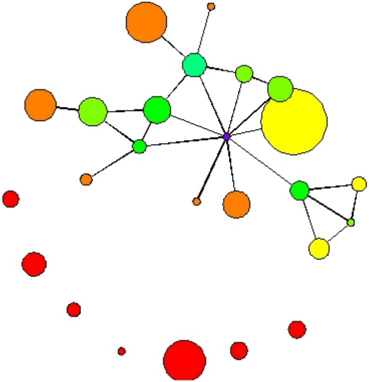
Correlation between DRS and Degree?



$\rho = 0.86$

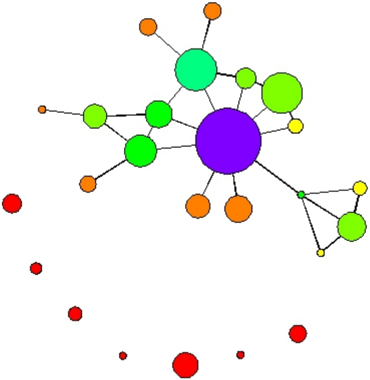
Correlation between DRS and Degree?



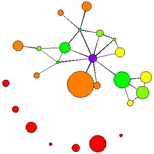


$\rho = 0.86$ $\rho = -0.07$

Correlation between DRS and Degree?

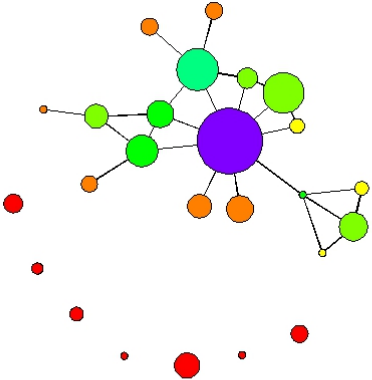


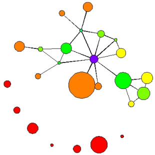
$\rho = 0.86$



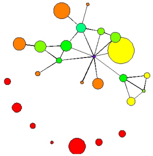
$\rho = -0.07$

Correlation between DRS and Degree?





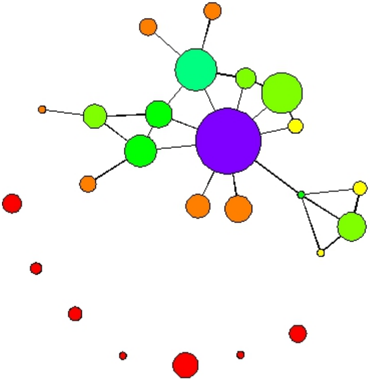
$\rho = -0.07$



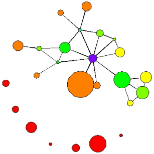
$\rho = -0.12$

$\rho = 0.86$

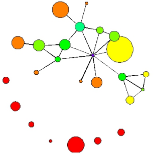
Correlation between DRS and Degree?



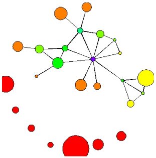
$\rho = 0.86$



$\rho = -0.07$

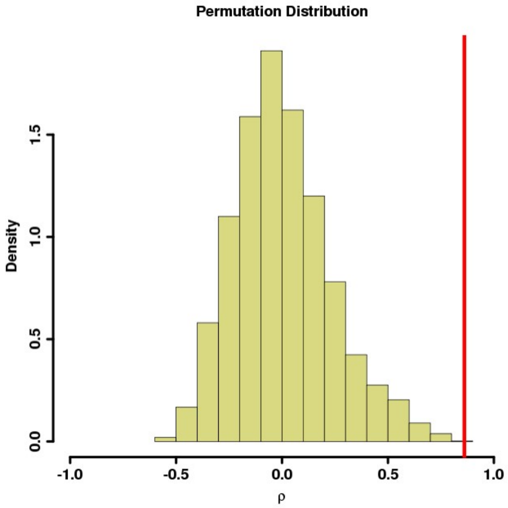


$\rho = -0.12$

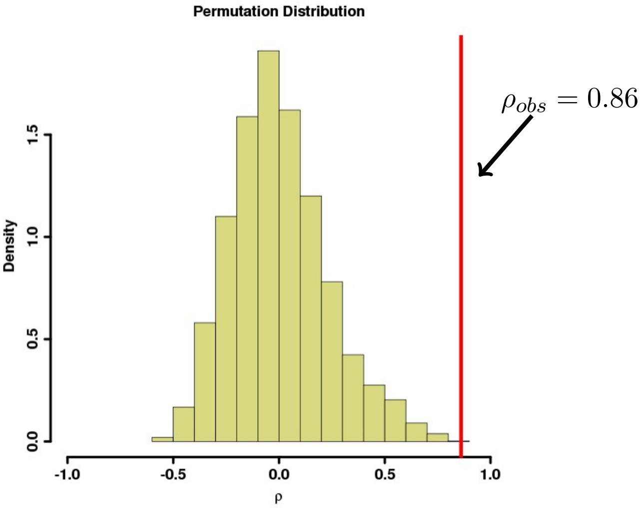


$\rho = -0.39$

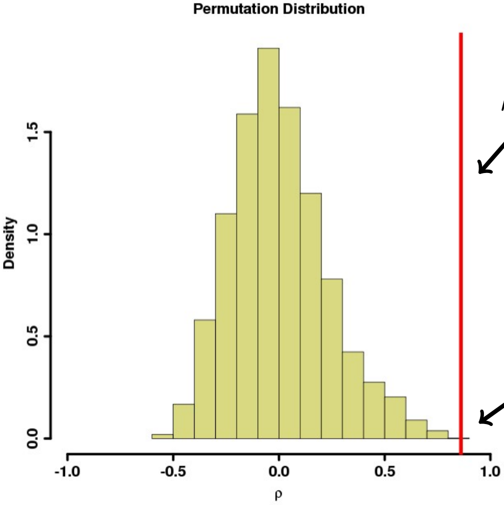
Correlation between DRS and Degree?



Correlation between DRS and Degree?



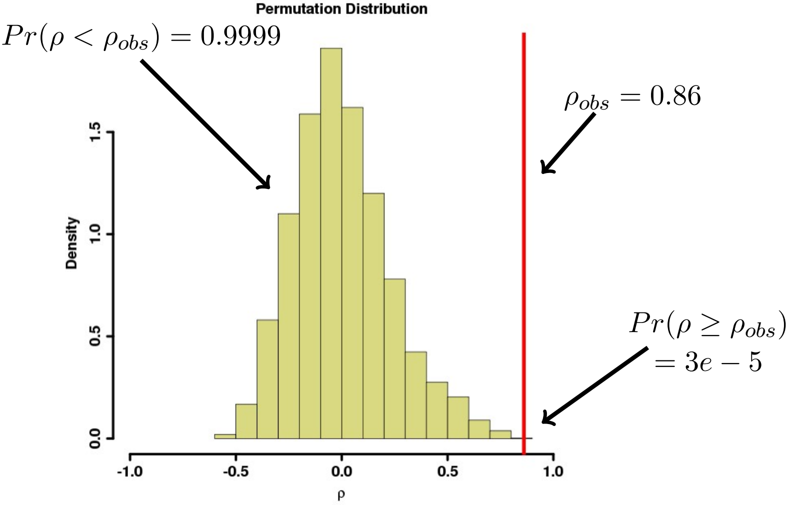
Correlation between DRS and Degree?



$\rho_{obs} = 0.86$

$Pr(\rho \geq \rho_{obs})$
$= 3e - 5$

Correlation between DRS and Degree?



Regression?

Regression?

- Can use Node Level Indices as independent variables in a regression

Regression?

- Can use Node Level Indices as independent variables in a regression
- Big assumption: ***position*** predicts the ***properties of those who hold them***

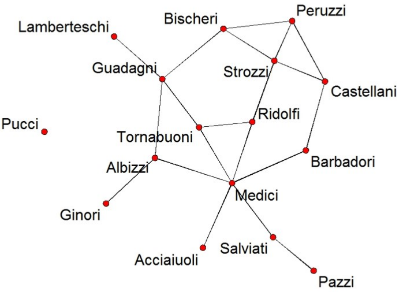
Regression?

- Can use Node Level Indices as independent variables in a regression
- Big assumption: ***position*** predicts the ***properties of those who hold them***
- Conditioning on NLI values, so dependence in accounted for ***assuming no error in the network***

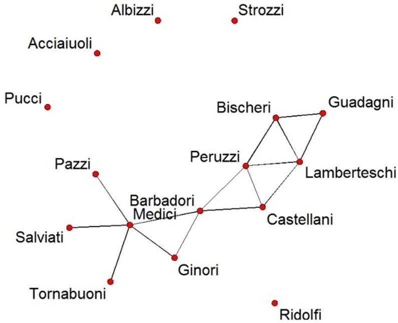
Regression?

- Can use Node Level Indices as independent variables in a regression
- Big assumption: ***position*** predicts the ***properties of those who hold them***
- Conditioning on NLI values, so dependence in accounted for ***assuming no error in the network***
- NLIs as dependent variables more problematic due to autocorrelation

Quadratic Assignment Procedure

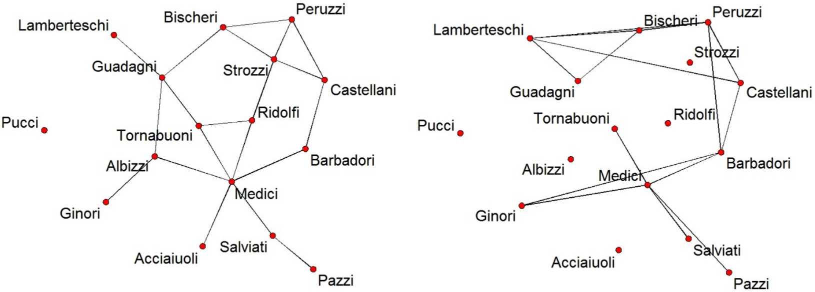


Marriage



Business

Quadratic Assignment Procedure



Marriage

Business

Graph Correlation

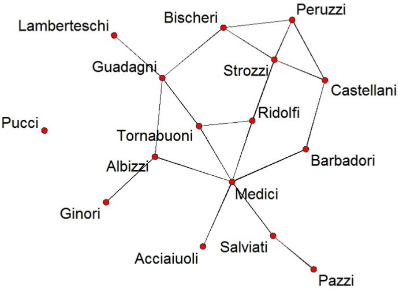
Graph Correlation

- Simple way of comparing graphs on the same vertex set by element

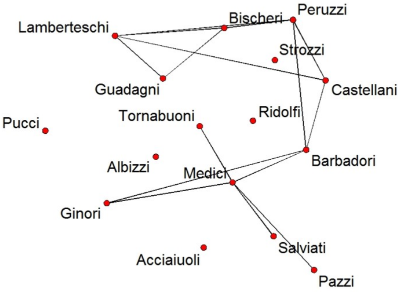
Graph Correlation

- Simple way of comparing graphs on the same vertex set by element
- $gcor\left(\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \right) = cor([1,1,1,0],[1,1,2,2])$

Do business ties coincide with marriages?



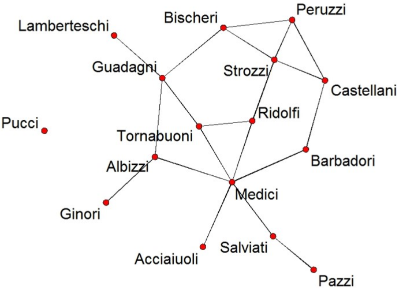
Marriage



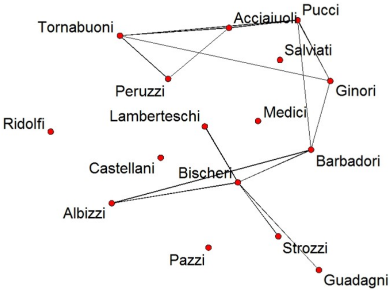
Business

$$\rho = 0.372$$

Do business ties coincide with marriages?

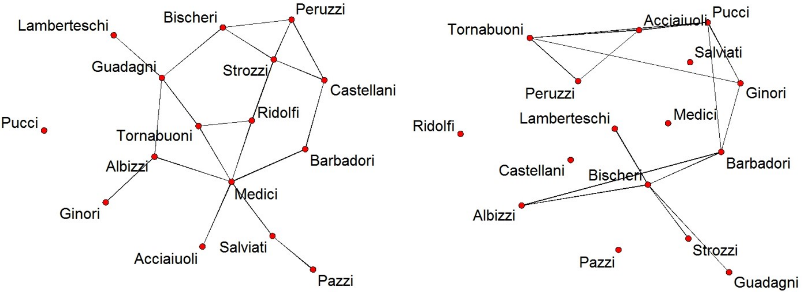


Marriage



Business

Do business ties coincide with marriages?

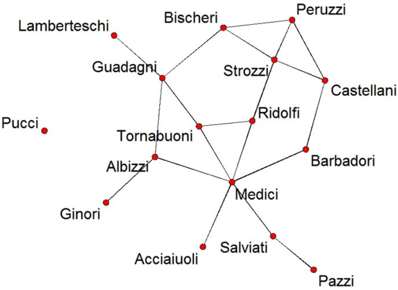


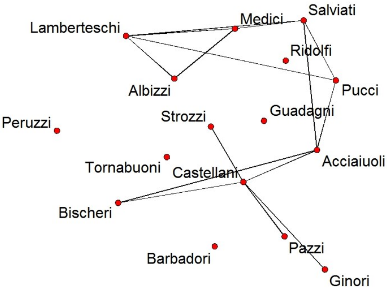
Marriage

Business

$$\rho = 0.169$$

Do business ties coincide with marriages?



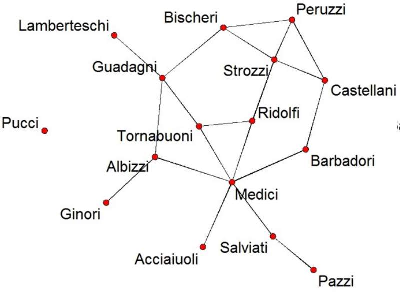


Marriage

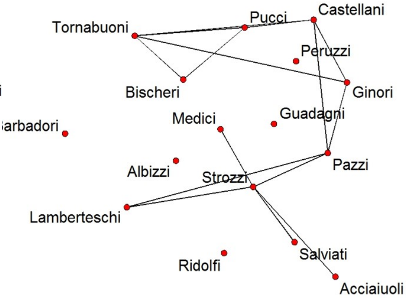
Business

$$\rho = -0.034$$

Do business ties coincide with marriages?



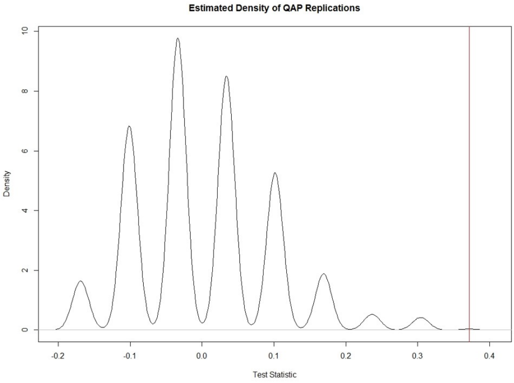
Marriage



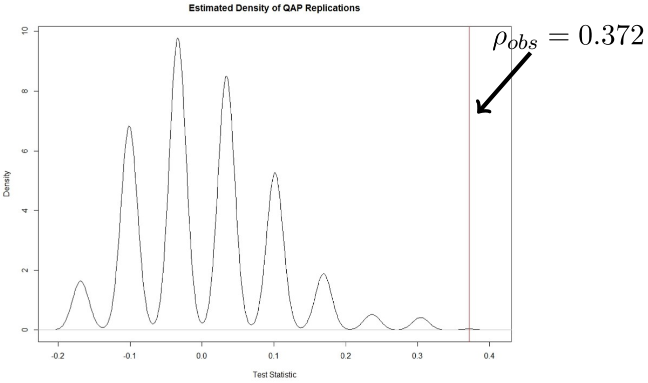
Business

$$\rho = -0.101$$

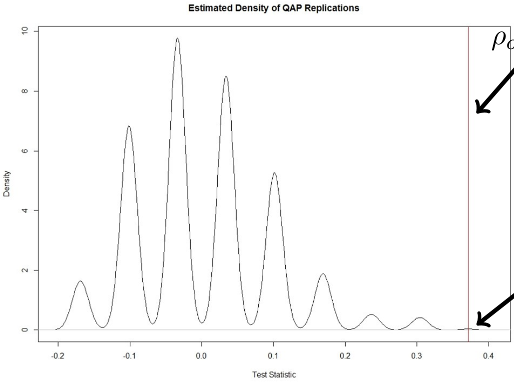
QAP Test


Estimated Density of QAP Replications
Density
Test Statistic
0
2
4
6
8
10
-0.2
-0.1
0.0
0.1
0.2
0.3
0.4

QAP Test


Estimated Density of QAP Replications
Density
Test Statistic
0
2
4
6
8
10
-0.2
-0.1
0.0
0.1
0.2
0.3
0.4
$\rho_{obs} = 0.372$

QAP Test

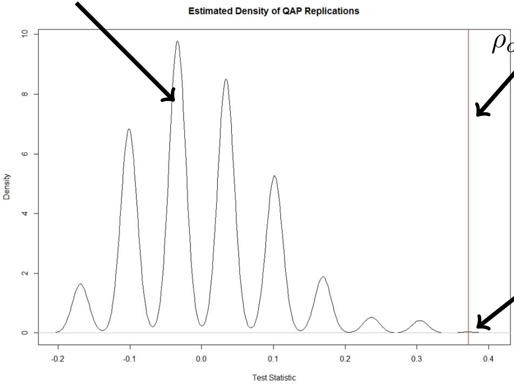


$\rho_{obs} = 0.372$

$Pr(\rho \geq \rho_{obs})$
$= .001$

QAP Test

$Pr(\rho < \rho_{obs}) = 0.999$



$\rho_{obs} = 0.372$

$Pr(\rho \geq \rho_{obs})$
$= .001$

Network Regression

Network Regression

- Family of models predicting social ties

Network Regression

- Family of models predicting social ties
  - Special case of standard OLS regression

Network Regression

- Family of models predicting social ties
  - Special case of standard OLS regression
  - Dependent variable is a network adjacency matrix

Network Regression

- Family of models predicting social ties
  - Special case of standard OLS regression
  - Dependent variable is a network adjacency matrix
- $\mathbf{E}Y_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \cdots + \beta_\rho X_{\rho ij}$

Network Regression

- Family of models predicting social ties
  - Special case of standard OLS regression
  - Dependent variable is a network adjacency matrix
- $\mathbf{E}Y_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \cdots + \beta_\rho X_{\rho ij}$
  - Where **E** is the expectation operator (analagous to "mean" or "average")

Network Regression

- Family of models predicting social ties
  - Special case of standard OLS regression
  - Dependent variable is a network adjacency matrix
- $\mathbf{E}Y_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \cdots + \beta_\rho X_{\rho ij}$
  - Where $\mathbf{E}$ is the expectation operator (analagous to "mean" or "average")
  - Y_{ij} is the value from i to j on the dependent relation with adjacency matrix Y

Network Regression

- Family of models predicting social ties
  - Special case of standard OLS regression
  - Dependent variable is a network adjacency matrix
- $\mathbf{E}Y_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \cdots + \beta_\rho X_{\rho ij}$
  - Where $\mathbf{E}$ is the expectation operator (analagous to "mean" or "average")
  - Y_{ij} is the value from i to j on the dependent relation with adjacency matrix Y
  - X_{kij} is the value of the kth predictor for the (i, j) ordered pair, and $\beta_0, \ldots \beta_\rho$ are coefficients

Data Prep

Data Prep

- Dependent variable is an adjacency matrix

Data Prep

- Dependent variable is an adjacency matrix
    - Standard case: dichotomous data

Data Prep

- Dependent variable is an adjacency matrix
  - Standard case: dichotomous data
  - Valued case

Data Prep

- Dependent variable is an adjacency matrix
  - Standard case: dichotomous data
  - Valued case
- Independent variables also in adjacency matrix form

Data Prep

- Dependent variable is an adjacency matrix
  - Standard case: dichotomous data
  - Valued case
- Independent variables also in adjacency matrix form
  - Always takes matrix form, but may be based on vector data

Data Prep

- Dependent variable is an adjacency matrix
    - Standard case: dichotomous data
    - Valued case
- Independent variables also in adjacency matrix form
    - Always takes matrix form, but may be based on vector data
    - eg. simple adjacency matrix, sender/receiver effects, attribute differences, elements held in common

Code Time

Sections 2.4-2.5

Baseline Models

Baseline Models

- treats social structure as maximally random given some fixed constraints

Baseline Models

- treats social structure as maximally random given some fixed constraints
- methodological premise from Mayhew

Baseline Models

- treats social structure as maximally random given some fixed constraints
- methodological premise from Mayhew
    - identify potentially constraining factors

Baseline Models

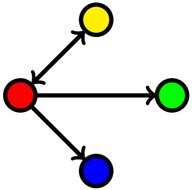
- treats social structure as maximally random given some fixed constraints
- methodological premise from Mayhew
  - identify potentially constraining factors
  - compare observed properties to baseline model

Baseline Models

- treats social structure as maximally random given some fixed constraints
- methodological premise from Mayhew
    - identify potentially constraining factors
    - compare observed properties to baseline model
    - useful even when baseline model is not ‘realistic’

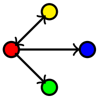
Types of Baseline Hypotheses

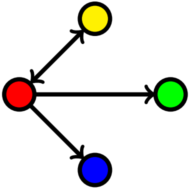
Types of Baseline Hypotheses



Empirical Network

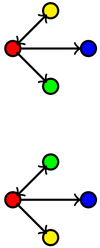
Types of Baseline Hypotheses

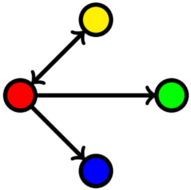




Empirical Network

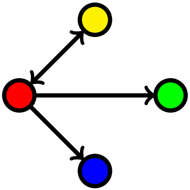
Types of Baseline Hypotheses



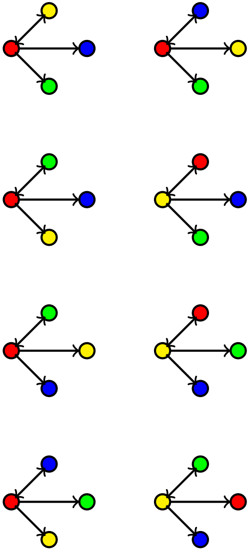


Empirical Network

Types of Baseline Hypotheses

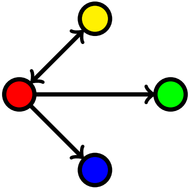


Empirical Network



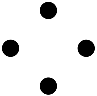
... etc

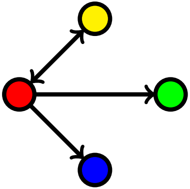
Types of Baseline Hypotheses



Empirical Network

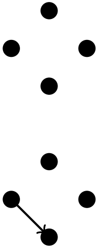
Types of Baseline Hypotheses

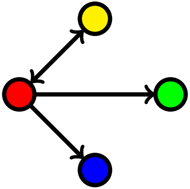




Empirical Network

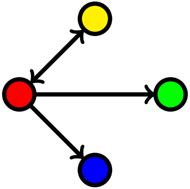
Types of Baseline Hypotheses



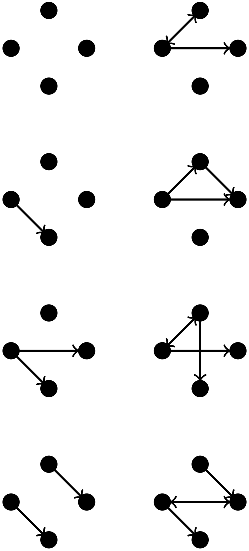


Empirical Network

Types of Baseline Hypotheses



Empirical Network



. . . etc

Types of Baseline Models

Types of Baseline Models

- **Size:** given the number of individuals, all structures are equally likely

Types of Baseline Models

- **Size:** given the number of individuals, all structures are equally likely
- **Number of edges/probability of an edge:** given the number of individuals and interactions (aka Erdös-Renyi random graphs)

Types of Baseline Models

- **Size:** given the number of individuals, all structures are equally likely
- **Number of edges/probability of an edge:** given the number of individuals and interactions (aka Erdös-Renyi random graphs)
- **Dyad census:** given number of individuals, mutuals, asymmetric, and null relationships

Types of Baseline Models

- **Size:** given the number of individuals, all structures are equally likely
- **Number of edges/probability of an edge:** given the number of individuals and interactions (aka Erdös-Renyi random graphs)
- **Dyad census:** given number of individuals, mutuals, asymmetric, and null relationships
- **Degree distribution:** given the number of individuals and each individual's outgoing/incoming ties

Types of Baseline Models

- **Size:** given the number of individuals, all structures are equally likely
- **Number of edges/probability of an edge:** given the number of individuals and interactions (aka Erdös-Renyi random graphs)
- **Dyad census:** given number of individuals, mutuals, asymmetric, and null relationships
- **Degree distribution:** given the number of individuals and each individual's outgoing/incoming ties
- **Number of triangles:** not implemented due to complexity – with ERGM, can condition on the *expected* number of triangles

Method

Method

- Select a test statistic (graph correlation, reciprocity, transitivity. . .)

Method

- Select a test statistic (graph correlation, reciprocity, transitivity. . .)
- Select a baseline hypothesis (what you're conditioning on)

Method

- Select a test statistic (graph correlation, reciprocity, transitivity. . .)
- Select a baseline hypothesis (what you're conditioning on)
- Simulate from the baseline hypothesis

Method

- Select a test statistic (graph correlation, reciprocity, transitivity. . .)
- Select a baseline hypothesis (what you're conditioning on)
- Simulate from the baseline hypothesis
- For each simulation, recalculate the test statistic

Method

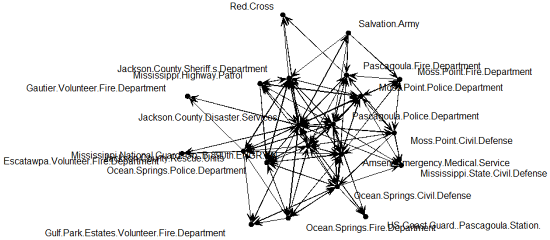
- Select a test statistic (graph correlation, reciprocity, transitivity…)
- Select a baseline hypothesis (what you're conditioning on)
- Simulate from the baseline hypothesis
- For each simulation, recalculate the test statistic
- Compare empirical value to null distribution, just as in standard statistical testing

Example

Transitivity in the Hurricane Frederic EMON

Example

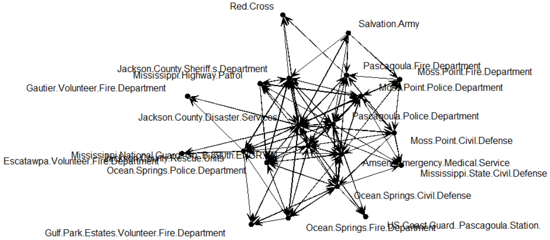
Transitivity in the Hurricane Frederic EMON



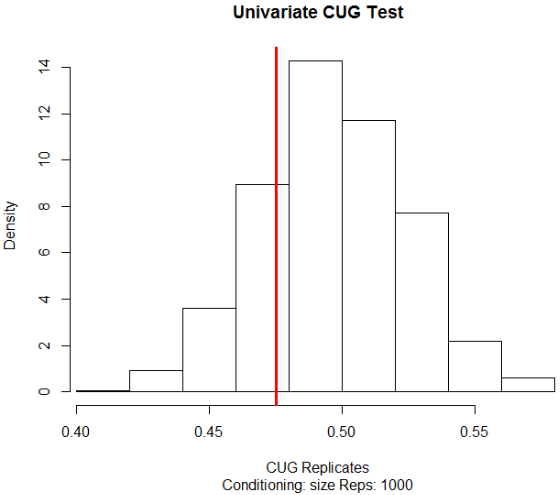
Example

Transitivity in the Hurricane Frederic EMON

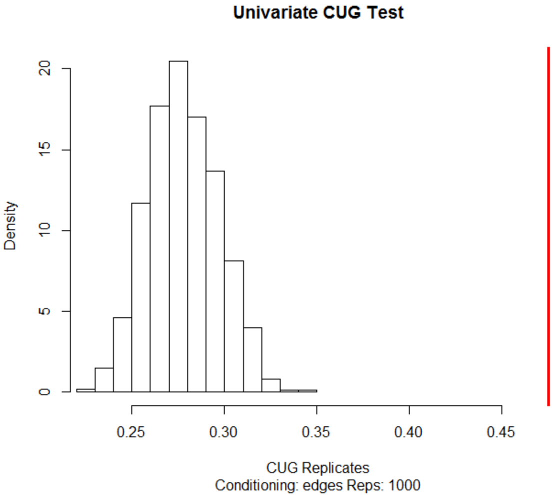
- $\rho = 0.475$
- indicates that roughly half the time that $i \to j \to k$, $i \to k$



Example


Univariate CUG Test
Density
0
2
4
6
8
10
12
14
0.40
0.45
0.50
0.55
CUG Replicates
Conditioning: size Reps: 1000

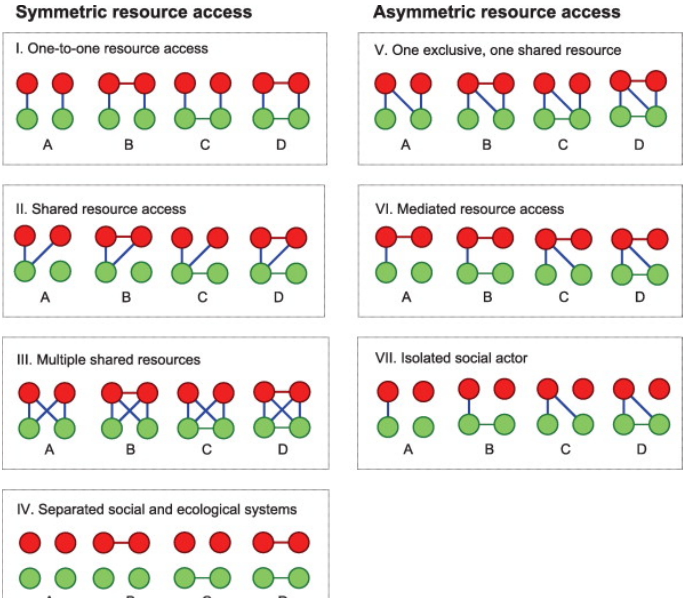
Example


Univariate CUG Test
Density
0
5
10
15
20
0.25
0.30
0.35
0.40
0.45
CUG Replicates
Conditioning: edges Reps: 1000

Bodin and Tengo

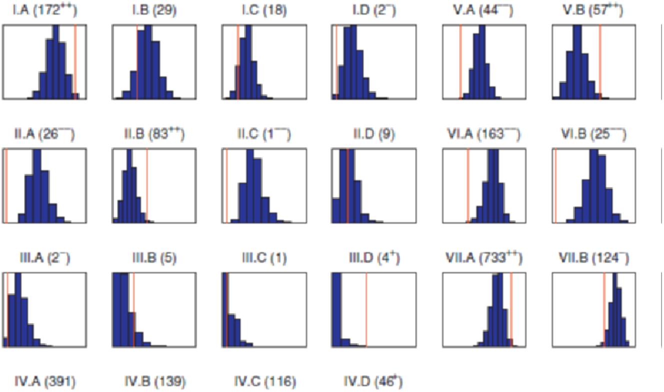
"Disentangling intangible social–ecological systems"

Bodin and Tengo



Bodin and Tengo

436 Ö. Bodin, M. Tengö / Global Environmental Change 22 (2012) 430–439



Caution!

Caution!

- Your selection of baseline model controls what hypothesis you're testing
- Changing the model can greatly change the results

Summary

Summary

- Network indices as independent variables in regression

Summary

- Network indices as independent variables in regression
- QAP regression (edges are the dependent variable)

Summary

- Network indices as independent variables in regression
- QAP regression (edges are the dependent variable)
- Network Autocorrelation Model (vertex attribute is dependent variable)

Summary

- Network indices as independent variables in regression
- QAP regression (edges are the dependent variable)
- Network Autocorrelation Model (vertex attribute is dependent variable)
- CUG tests (network is dependent variable)