

Slide 1

## Working with Missing Data

Steve Borgatti  
LINKS Center Workshop on Social  
Network Analysis

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

## The problem:

- Some respondents did not participate in the survey, leaving blank rows in the network data matrix
- Important note: if you enter data as edgelist or nodelist using dl file, missing values are automatically converted to zeros.
  - How would you convert them back?

		1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6
	A	A	B	B	C	G	G	L	M	P	P	P	R	S	S	T	
1	ACCIAIUOL	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
2	ALBIZZI	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0
3	BARBADORI	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
4	BISCHERI																
5	CASTELLAN	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0
6	GINORI	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	GUADAGNI	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	1
8	LAMBERTES	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
9	MEDICI	1	1	1	0	0	0	0	0	0	0	0	1	1	0	1	
10	PAZZI																
11	PERUZZI	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0
12	PUCCI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	RIDOLFI	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	
14	SALVIATI	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
15	STROZZI	0	0	0	1	1	0	0	0	0	0	1	0	1	0	0	0
16	TORNABUON	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

Size of the problem

- Counting the number of missing values with Tools | Freq.
  - Select "matrices"

Files

Input dataset: padgwithessing

Output dataset: padgwithessing

Compute Frequencies For:  
 Rows  
 Columns  
 **Matrices**

Output:  
 Frequencies  
 **Percentages**

Display of squares:  
 Exclude  
 **Include**

	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6
A	B	C	G	L	M	P	R	S	T							
1 ACCIAIUOL	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
2 ALBIZZI	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0
3 BARBADORI	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
4 BISCHERI																
5 CASTELLAN	0	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0
6 GINORI	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7 GUADAGNI	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1
8 LAMBERTES	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
9 MEDICI	1	1	1	0	0	0	0	0	0	0	0	0	1	1	0	1
10 PAZZI																
11 PERUZZI	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1
12 PUCCI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13 RIDOLEI	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
14 SALVIATI	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
15 STROZZI	0	0	0	1	1	0	0	0	0	1	0	1	0	0	0	0
16 TORNABUON	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0

Output:

	1
	-----
0,000	0,725
1,000	0,150
blank	0,125

## Standard Solutions

- Convert missings to zeros (since you did NOT observe a tie)
  - Re-run having converted missings to ones, to see how different the results could be
- Convert missings to zeros and ones at random, using density of the matrix as guide
- Impute the missing values using other information
  - Symmetricity
  - QAP regression

---

---

---

---

---

---

---

---

---

---

---

---

---

## One solution

- Suppose the data are largely symmetric, and the social relation is logically symmetric
  - Marriage to
  - Saw movie with
- Then we can impute the missing data from the transpose of the matrix
  - i.e., assume that if A says B is a friend, then if B had participated, s/he would have said A was a friend too
  - So, fill in missing row with the corresponding column

	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6
A	A	B	C	G	L	M	P	P	R	S	S	T	-	-	-
1 ACCIAIUOL	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
2 ALBIZZI	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0
3 BARBADORI	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
4 BISCHERI															
5 CASTELLAN	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1
6 GINORI	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
7 GUADAGNI	0	1	0	1	0	0	0	1	0	0	0	0	0	0	1
8 LAMBERTES	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
9 MEDICI	1	1	1	0	0	0	0	0	0	0	0	1	1	0	1
10 PAZZI															
11 PERUZZI	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1
12 PUCCI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13 RIDOLFI	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
14 SALVIATI	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
15 STROZZI	0	0	1	1	0	0	0	0	0	1	0	1	0	0	0
16 TORNABUON	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

## Doing it ...

- In matrix algebra, you can do it with the REPLACENA command:
  - `newdata = replacena(olddata transp(olddata))`
- Syntax
  - `> <newds> = replacena(<ds1> <ds2>)`
  - Where ds1 is the dataset that contains missing values and ds2 is the dataset from which to draw the correct values

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

## What if the relation is not symmetric?

- If the network with missing data is directed (i.e., not necessarily symmetric), we can't do this trick.
  - E.g., you have asked “who do you get advice from”
- Unless ...
  - Suppose you have asked both “who do you get advice from” and “who gets advice from you”
  - We can use one to fill in data for the other

---

---

---

---

---

---

---

---

---

---

---

---

---

**Example (without missing data)**

Give Info To

	A	B	C	D
A	0	1	0	1
B	1	0	1	0
C	0	1	0	1
D	0	1	0	1

Get Info From

	A	B	C	D
A	0	1	0	0
B	1	0	1	1
C	0	1	0	0
D	1	0	1	1

TRANSCOPE(Get Info From)

	A	B	C	D
A	0	1	0	1
B	1	0	1	0
C	0	1	0	1
D	0	1	0	1

Should be the same  
So if a row is missing in GiveInfo, we should be able to fill in from row in transpose of GetInfo

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

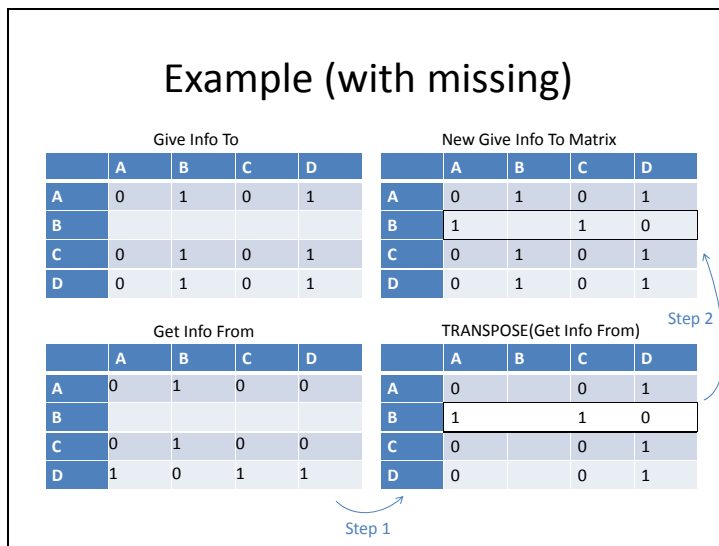
---

---

---

---






---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

**Do it in UCINET using “REPLACENA”**  
(in tools|matrix algebra)

- Example
  - > InfoTo = replacena(giveinfo transp(getinfo))
  - > friends = replacena(rawfriends transp(rawfriends))
- Syntax
  - > <newds> = replacena(<ds1> <ds2>)
  - Where ds1 is the dataset that contains missing values and ds2 is the dataset from which to draw the correct values

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---