SOCIAL NETWORK ANALYSIS for DATA SCIENTISTS

today's menu: LAB: ERGM I (LECTURE Week 5)

Your lecturer: Claudia

Playdate: September, 30th, 2025

Menu'

- Homeplay Discussion

- Markovian Terms

- Exogenous Terms

Homeplay Discussion

Q1 -explore the network

Explore the network (N, E, attributes, D or U) and run the necessary descriptive statistics (remember to check the class and type of each attribute).

```
  snafun::g_summary(Hookups)


## $number_of_vertices
## [1] 44
##
## $number_of_edges
## [1] 46
##
## $density
## [1] 0.049
##
## $reciprocity
## [1] 1
##
## $transitivity
## [1] 0
##
## $mean_distance
## [1] Inf
```

Q2: What are the hypotheses formulated in the book for this example?

Q2: What are the hypotheses formulated in the book for this example?

1. Characters hookup with other characters of similar age

2. Characters hookup with other characters in the same job position

3. Characters hookup with other characters of the same race

4. Characters hookup with other characters of different sex

5. Younger characters have more hookups

6. Characters in the show for a short time had less opportunity for hookups

Q2: What are the hypotheses formulated in the book for this example?

1. Characters hookup with other characters of similar age absdiff

2. Characters hookup with other characters in the same job position nodematch

3. Characters hookup with other characters of the same race nodematch

4. Characters hookup with other characters of different sex nodematch

5. Younger characters have more hookups nodecov

6. Characters in the show for a short time had less opportunity for hookups nodecov

Q3

Open page 73 and insert the terms specified in the example in a series of nested models. The model on page 73 is your final complete model—nest at least 3 of them to explore the effects of the terms one by one.

3a) which is the best model?

3b) Can you tell whether each model is dyadic (inter)dependent or independent?

3c) Is each of these variables exogenous or endogenous?

3d) Comment on the results. What is the substantive meaning of each of these terms? Are the hypotheses supported? Comment on the odd ratios too.

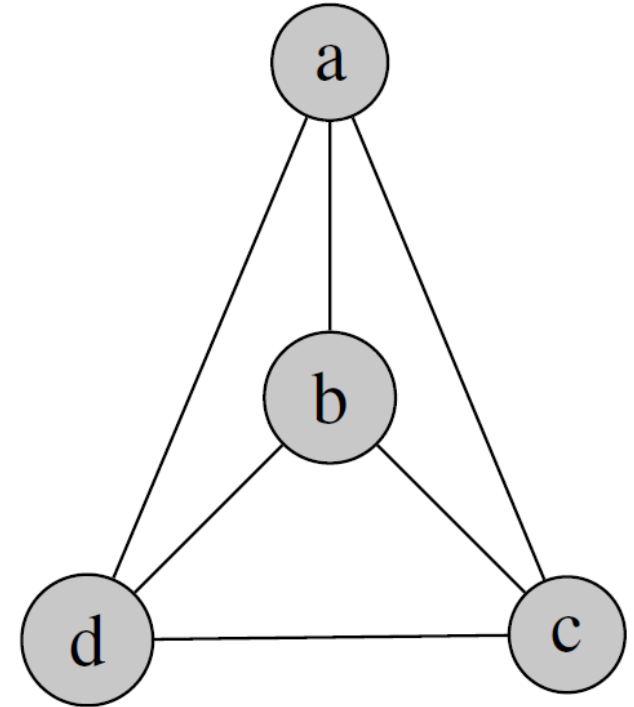3e) What is the difference between absdif and nodecov estimated on birthyear?

Q4

Formulate one additional hypothesis on the network's structure and then nest another model, adding one variable to explore the network's structure and test this final hypothesis.

Markovian Terms

Markovian effects

Effects in a Markov graph

- It is hypothesized that only neighboring nodes impact the odds of a link

- This level of dependence is also referred to as a nearest-neighbors level of dependence

- or dyadic dependence.

- we need to find all structures that are subject to a Markov (or nearest-neighbor) level of dependence

- Key words: Markovian, structural, dyadic dependent, structure of neighboring nodes
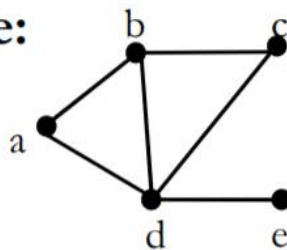
Undirected structural/Markovian/endogenous/dyadic dependent terms

http://math.iisc.ernet.in/~imi/downloads/NBSSE/Noshir%20Contractor/Lecture%2007.pdf

Terms for undirected structural
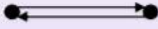
Probability that that number of the observed undirected pattern is statistically significant

- `edges` - probability that the number of observed edges is statistically significant.

- `kstar(2)` - probability that the number of observed 2-stars is statistically significant.

- `kstar(3)` - probability that the number of observed 3-stars is statistically significant.

- `kstar(K)` - probability that the number of observed K-stars is statistically significant.

- `triangles` - probability that the number of observed triangles is statistically significant.

Directed structural/Markovian/endogenous/dyadic dependent terms

http://math.iisc.ernet.in/~imi/downloads/NBSSE/Noshir%20Contractor/Lecture%2007.pdf

Terms for directed structural

Probability that that number of the observed directed pattern is statistically significant

- `asymmetric` <- arc - Probability that that number of the observed asymmetric dyads is statistically significant

- `mutual` <- reciprocity - Probability that that number of the observed mutual dyads is statistically significant

- `istar(2)` <- 2 in star - Probability that that number of the observed in 2 two stars is statistically significant

- `ostar(2)` <- 2 out star - Probability that that number of the observed out 2 two stars is statistically significant

- `m2star` <- mixed 2 star - Probability that that number of the observed mixed 2 two stars is statistically significant

- `istar(k)` <- k in star - Probability that that number of the observed in K two stars is statistically significant

- `ostar(k)` <- k out star - Probability that that number of the observed out k two stars is statistically significant

Edgewise and dyadwise shared partners

(a) 1 partner   (b) 2 partners   (c) 3 partners
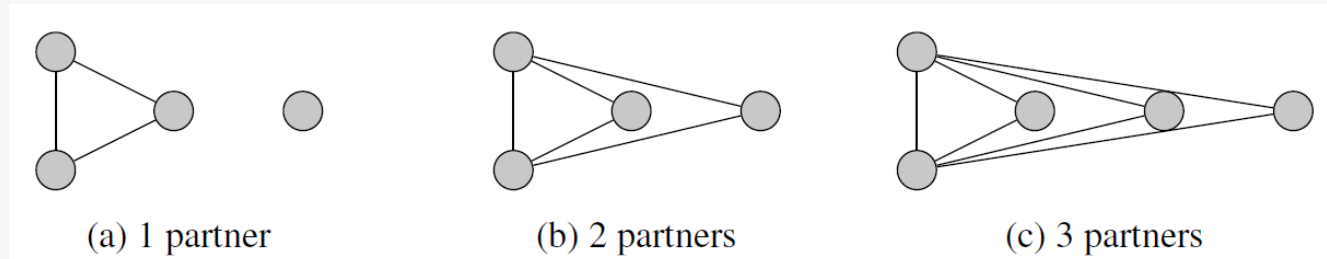
Figure 15: Shared partners

(a) 1 partner   (b) 2 partners   (c) 3 partners
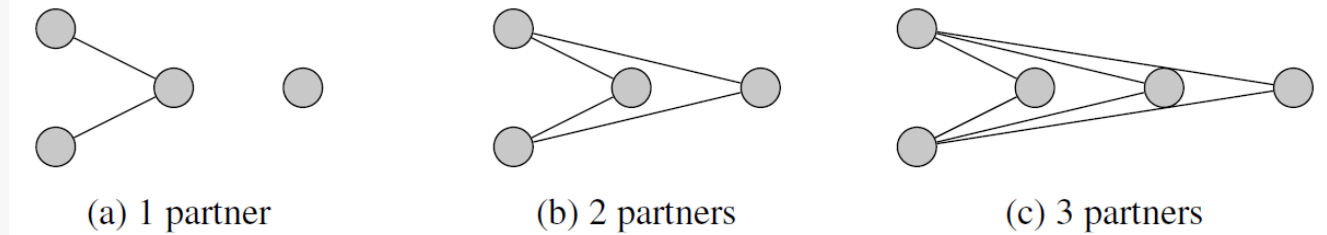
Figure 16: Shared partners (dyadic)

Terms for shared partners

Edgewise shared partners

- `esp` - undirected

- `desp` - directed (5 options) 1) outgoing two path, 2) incoming two path, 3) reciprocated 2 path 4) outgoing shared partner, 5) incoming shared partner.

Absence of connections in a structure where these connections would be possible

- `nsp` - undirected edgewise NON-shared partners

- `dnsp` - directed edgewise NON-shared partners.

Dyadwise shared partners

- `dsp` - undirected dyadwise shared partner

- `ddsp` - directed dyadwise shared partner

More about triangles: the triad census

3 digits [mutual][asymmetric][no-edge]

Local Structure – Transitivity

Davis and Leinhardt (1972)
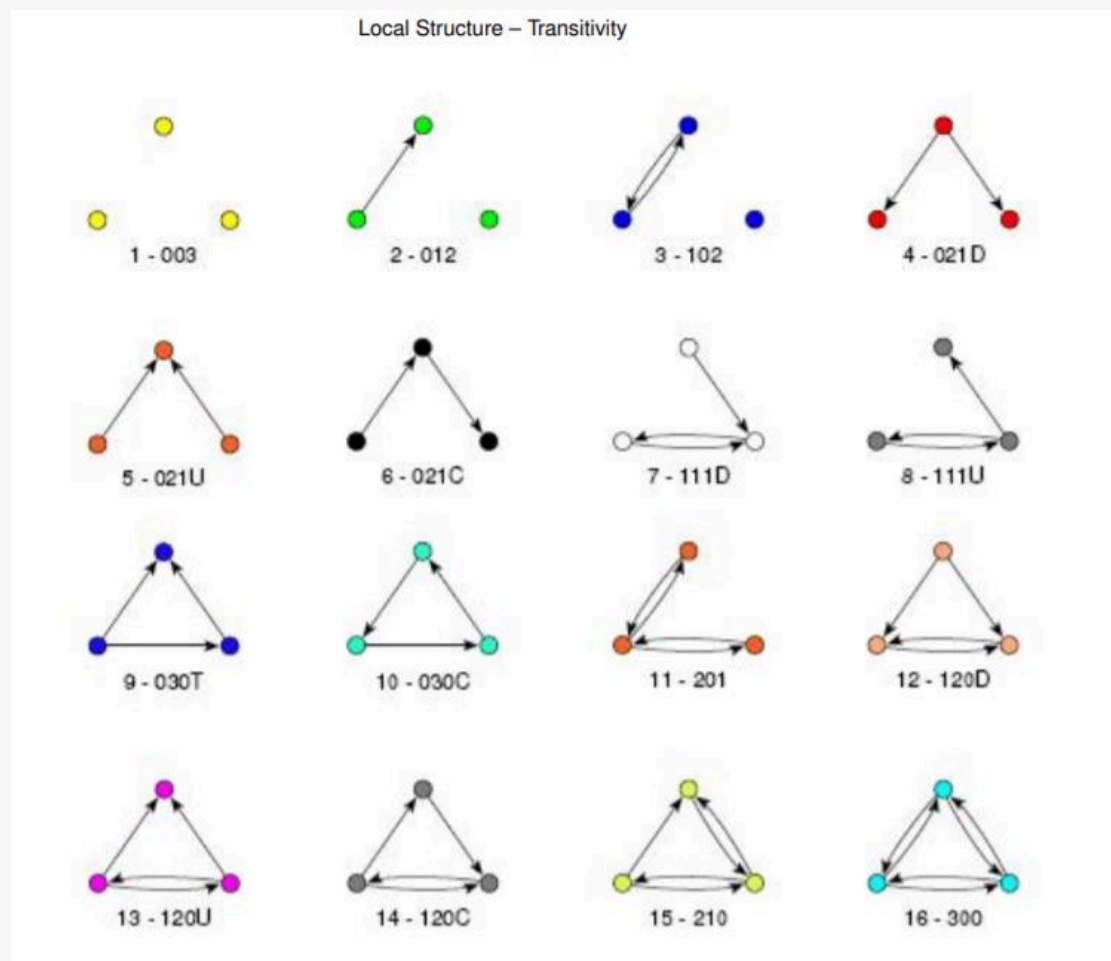
undirected networks

4 options (no edge or reciprocated edges)

- 003, 102, 201, 300

Local Structure – Transitivity
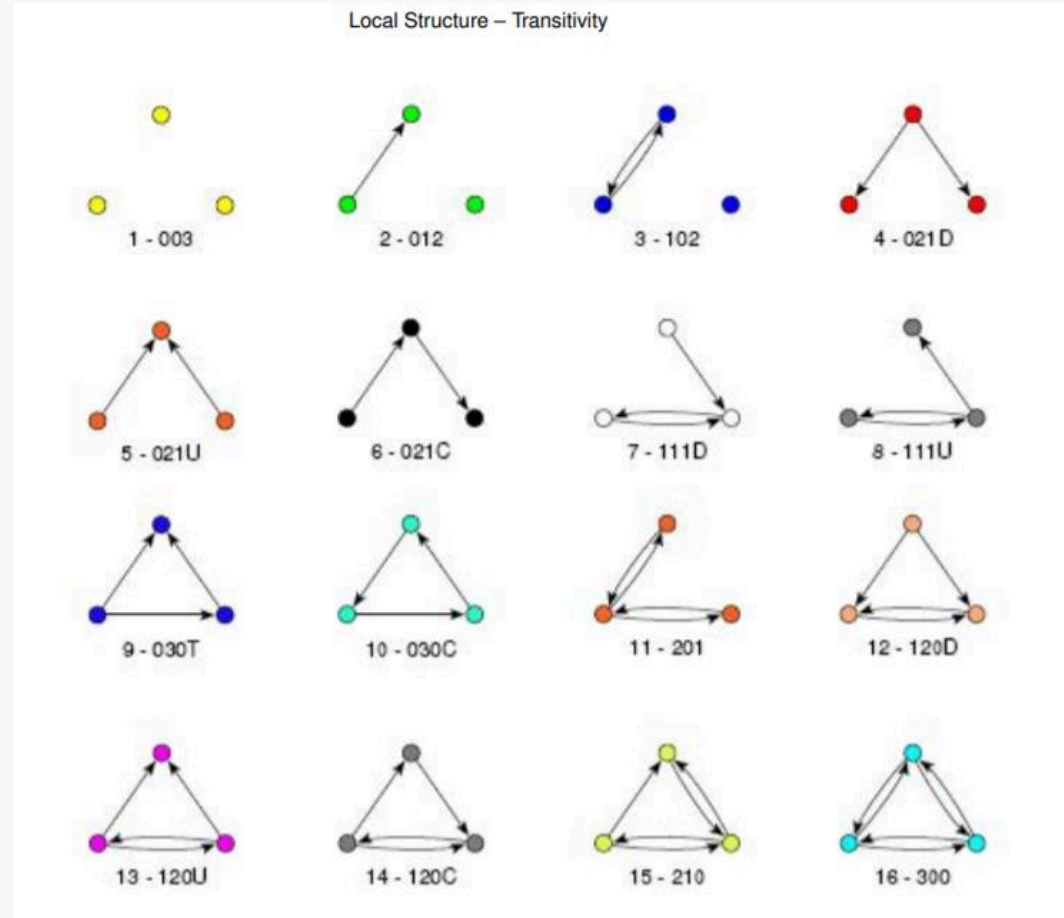
Terms for triads

- `triadcensus` - inserts all the 16 options in the model (rarely a good idea)

Local Structure – Transitivity

1 - 003   2 - 012   3 - 102   4 - 021D
5 - 021U   6 - 021C   7 - 111D   8 - 111U
9 - 030T   10 - 030C   11 - 201   12 - 120D
13 - 120U   14 - 120C   15 - 210   16 - 300
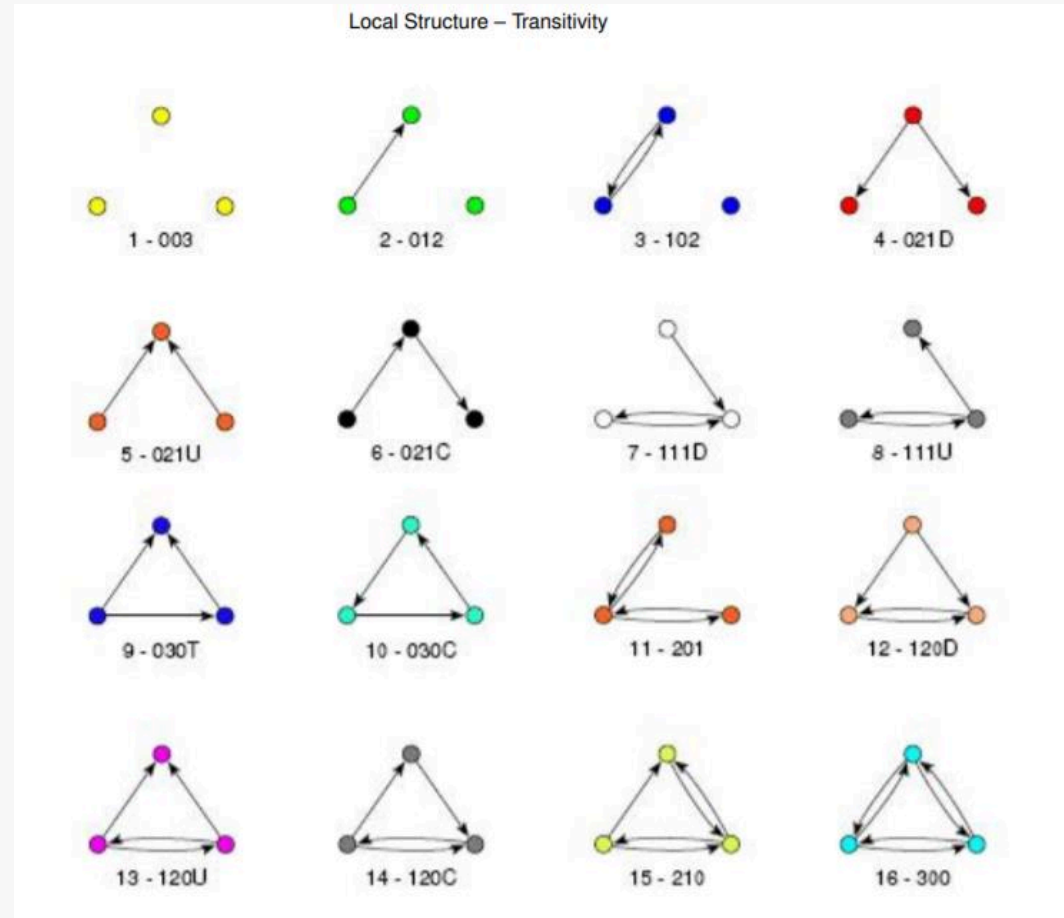
Terms for triads

- `ctriple` - 030C - Cyclic triples: defined as a set of edges of the form (i→j), (j→k), (k→i)

Local Structure – Transitivity

1 - 003   2 - 012   3 - 102   4 - 021D

5 - 021U   6 - 021C   7 - 111D   8 - 111U

9 - 030T   10 - 030C   11 - 201   12 - 120D

13 - 120U   14 - 120C   15 - 210   16 - 300
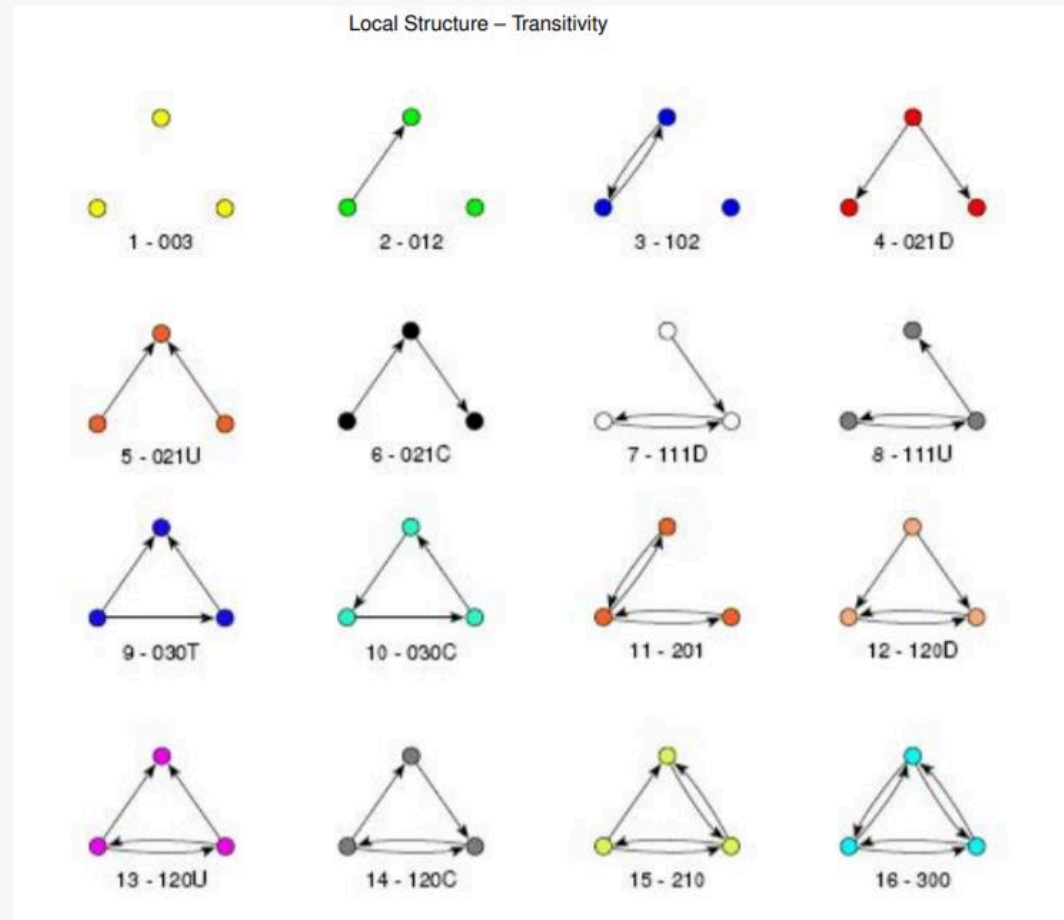
Terms for triads

- `ttriple` - O30T - transitive triples: defined as a set of edges (i→j), (j→k), (i→k)

Local Structure – Transitivity

1 - 003 2 - 012 3 - 102 4 - 021D
5 - 021U 6 - 021C 7 - 111D 8 - 111U
9 - 030T 10 - 030C 11 - 201 12 - 120D
13 - 120U 14 - 120C 15 - 210 16 - 300
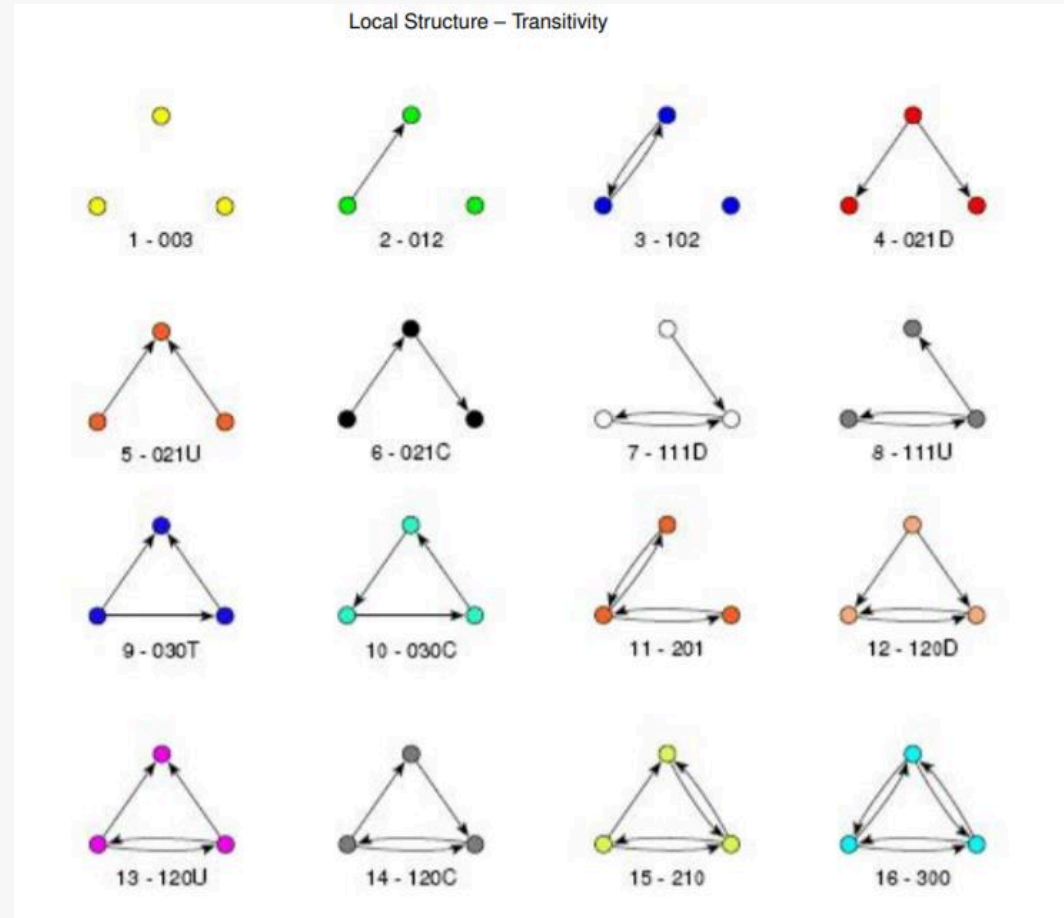
Terms for triads

- `intransitive` - 111D, 201, 111U, 021C - Intransitive triads

Local Structure – Transitivity
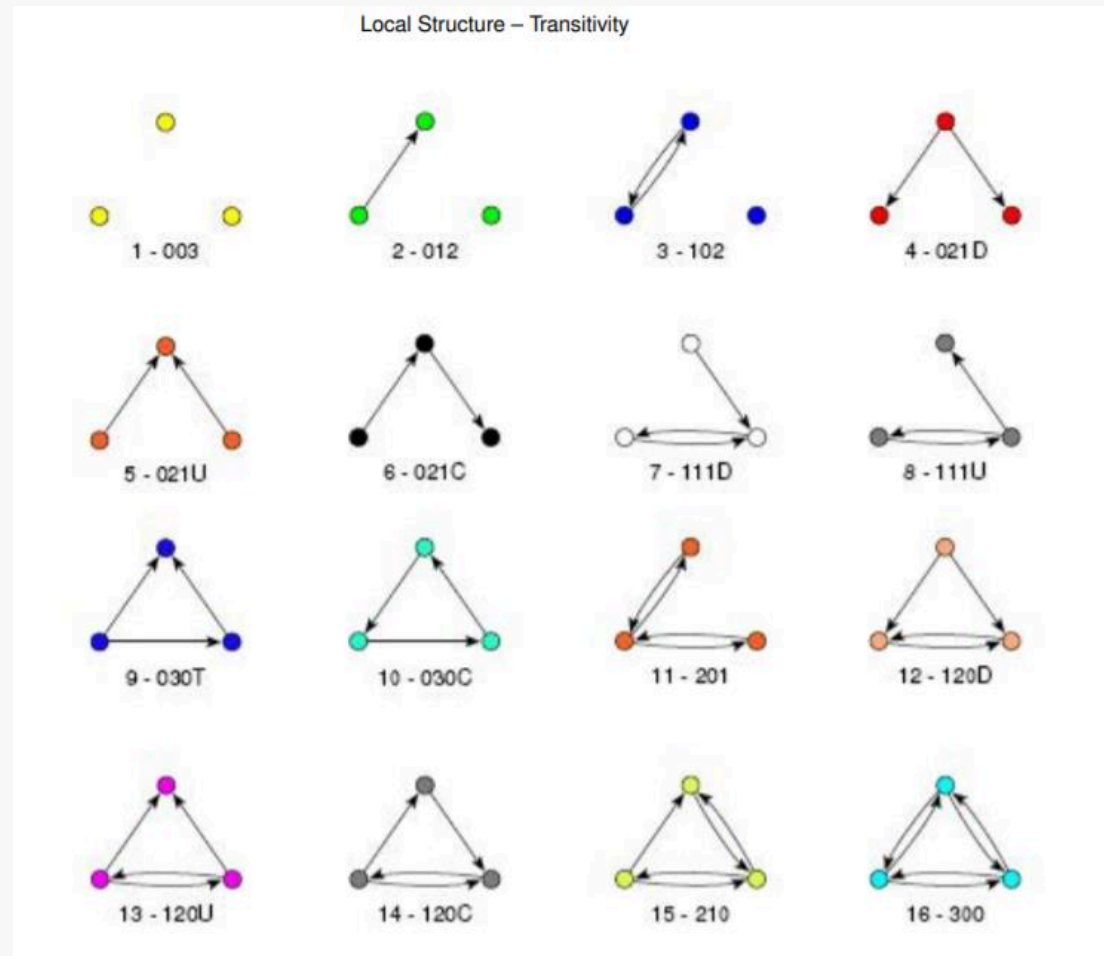
Terms for triads

- `transitive` - 120D, 120U, 300, also 030T, 030C - Transitive triads

Local Structure – Transitivity

1 - 003　　2 - 012　　3 - 102　　4 - 021D
5 - 021U　　6 - 021C　　7 - 111D　　8 - 111U
9 - 030T　　10 - 030C　　11 - 201　　12 - 120D
13 - 120U　　14 - 120C　　15 - 210　　16 - 300
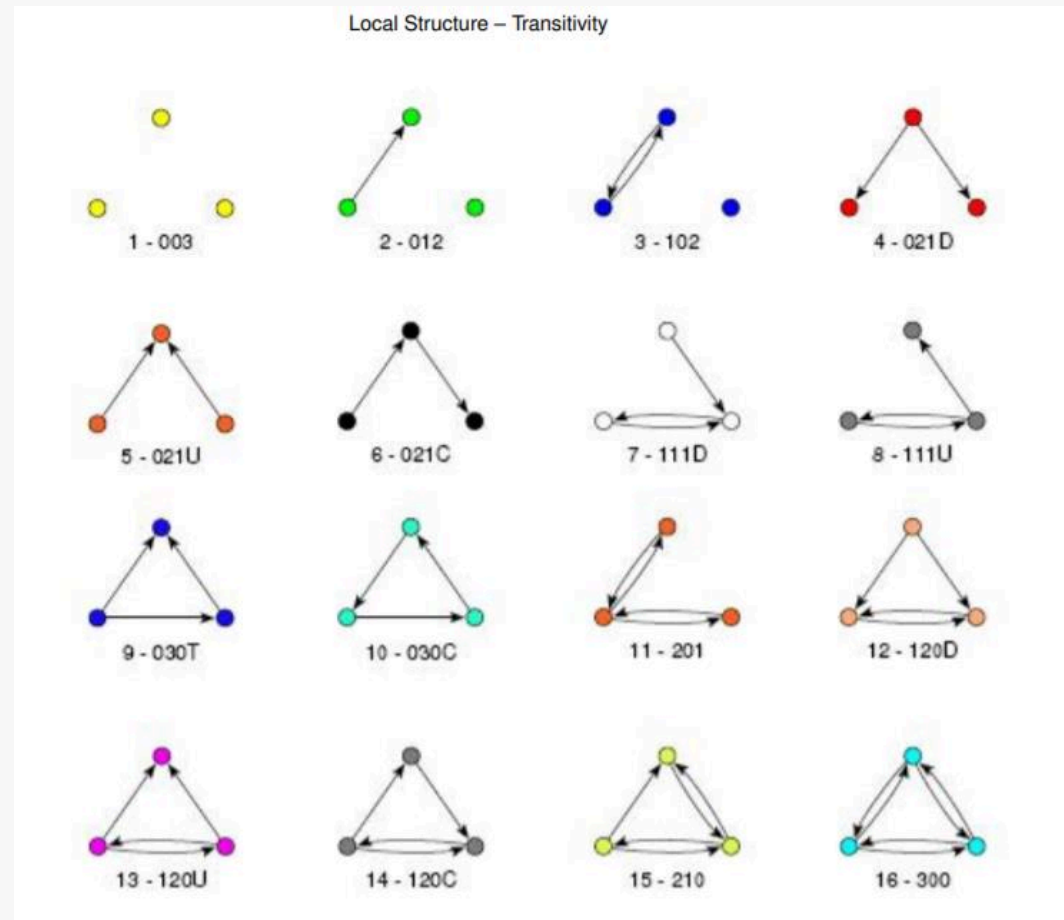
Terms for triads

- `balance` - 102, 300

Local Structure – Transitivity

1 - 003, 2 - 012, 3 - 102, 4 - 021D, 5 - 021U, 6 - 021C, 7 - 111D, 8 - 111U, 9 - 030T, 10 - 030C, 11 - 201, 12 - 120D, 13 - 120U, 14 - 120C, 15 - 210, 16 - 300
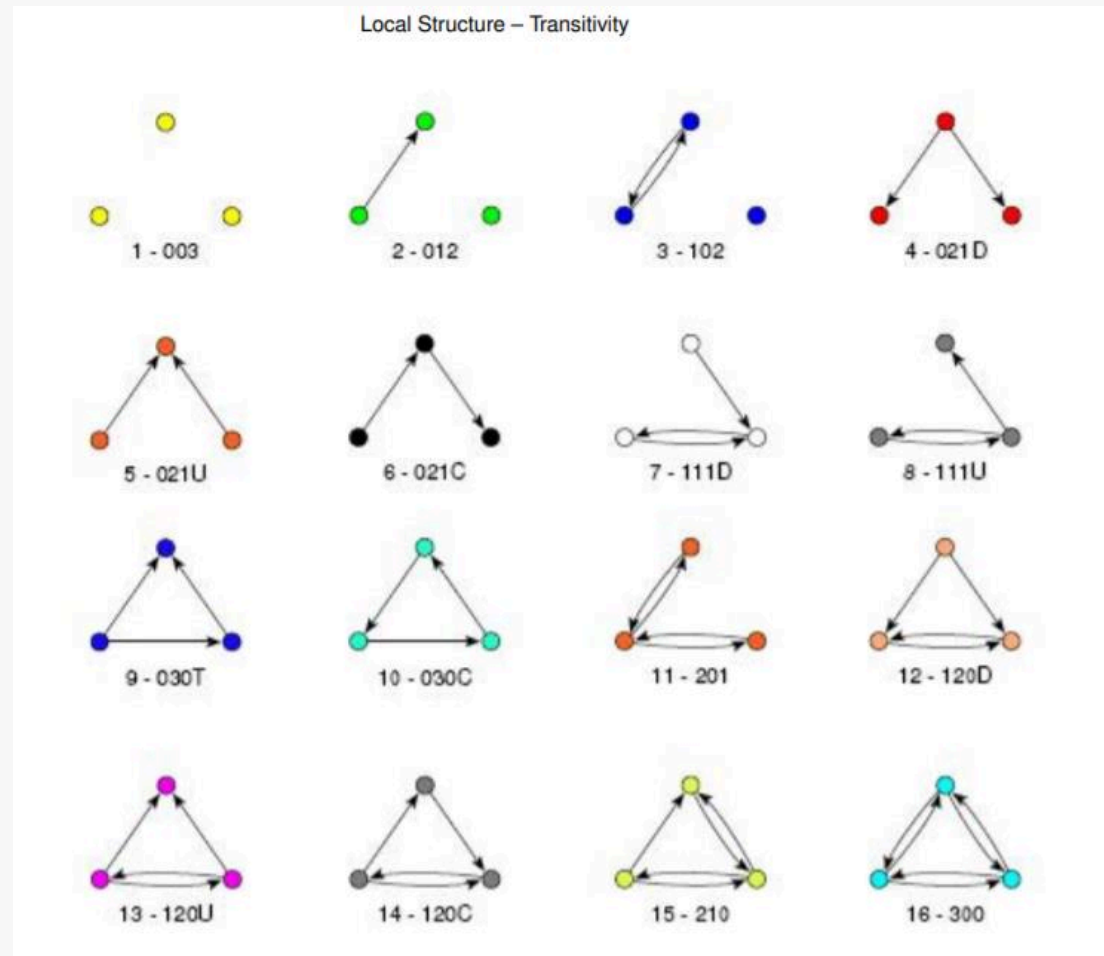
Cycles

`cycle(k)` <- k can be any number.

`cycle(3) == ctriple`
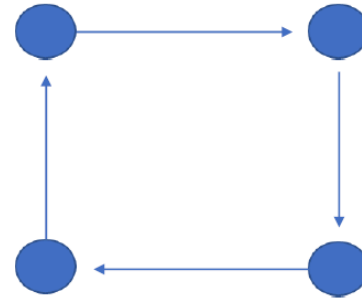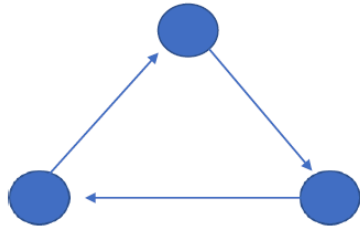
Exogenous Terms

Extent to which the attribute values influence edge formation so that it is non-random under that condition.

- nodecov - undirected - numeric var

- nodeicov - directed in - numeric var

- nodeocov - directed out - numeric var

Florentine family example:

The extent to which being wealthier increases the probability of getting married.

same as in a logit model

The extent to which being in a certain category influences edge formation so that tie formation is non-random under that condition

- `nodefactor` - undirected - categorical var

- `nodeifactor` - directed in - categorical var

- `nodeofactor` - directed out - categorical var

same as in a logit model

Extent to which common features measured in terms of distance similarity influence edge formation, so that edge formation is non-random under that condition.

`absdiff` - numeric var

Florentine family example:

The extent to which being equally wealthy increases the probability of getting married.

- poor marry poor

- rich marry rich

This term implicitly attributes a weight to edges since the two connected nodes share the absolute difference

For this reason it is inserted in the model as a matrix rather than as a vector

(even if you don't see this happening)

Extent to which nodes characterized by a specific category belonging to a certain attribute form ties with other node characterized by the same category, so that tie formation under that condition is non-random

`nodematch` - categorical var

This term measures homophily. The probability of a tie due to similarity

- If the coefficient θ estimated in the model is negative it is measuring heterophily: the probability of a tie due to differences.

Examples:

- friends in primary school, it is likely that girls are friend with girls (homophily due to sex)

- heterosexual romantic relationships a tie is due to heterophily in sex

Extent to which nodes denoted by different categories of an attribute form ties, so that tie formation under these circumstances is non-random.

`nodemix` - disassortative mixing - categorical var

This term measures heterophily. The probability of a tie due to dissimilarity

- If the coefficient θ estimated in the model is negative it is measuring homophily: the probability of a tie due to similarity.

Same example as the previous slide

Extent to which the ties formed in another context influence tie formation in the context of the current model, so that tie formation under that circumstances is non-random.

`edgecov` - Matrix attribute.

It can be used to insert edge attributes in the model. They need to be specified as a matrix

E.g., how often two students study together can be inserted in the model as a weighted matrix
NOTE: using `edgecov` you can run a QAP logistic model using ERGMs!

Cross tabulation between edge formation and attributes of interest

Mixing matrix. THIS IS NOT A MODEL TERM

```
snafun::make_mixingmatrix()
```

```
snafun::make_mixingmatrix(floM, "Wealth")
```

```
##        from
## to      3  8 10 20 27 32 36 42 44 48 49 55 103 146 Sum
##   3     0  0  0  0  0  0  0  0  0  0  0  0   0   0   0
##   8     0  0  0  0  0  0  1  1  1  1  0  0   0   0   4
##   10    0  0  0  0  0  0  0  0  0  1  0  0   2   0   3
##   20    0  0  0  0  0  0  0  0  0  0  1  1   0   1   3
##   27    0  0  0  0  0  0  0  0  0  1  0  0   1   1   3
##   32    0  0  0  0  0  0  1  0  0  0  0  0   0   0   1
##   36    0  1  0  0  0  1  0  0  0  0  0  0   1   0   3
##   42    0  1  0  0  0  0  0  0  0  0  0  0   0   0   1
##   44    0  1  0  0  0  0  0  0  0  0  1  0   0   1   3
##   48    0  1  1  0  1  0  0  0  0  0  0  0   1   0   4
##   49    0  0  0  1  0  0  0  0  1  0  0  0   0   1   3
##   55    0  0  0  1  0  0  0  0  0  0  0  0   1   0   2
##   103   0  0  2  0  1  0  1  0  0  1  0  1   0   0   6
##   146   0  0  0  1  1  0  0  0  1  0  1  0   0   0   4
##   Sum   0  4  3  3  3  1  3  1  3  4  3  2   6   4  20
```

BYe ByE