# Mixing pattern quantification in node-attributed networks

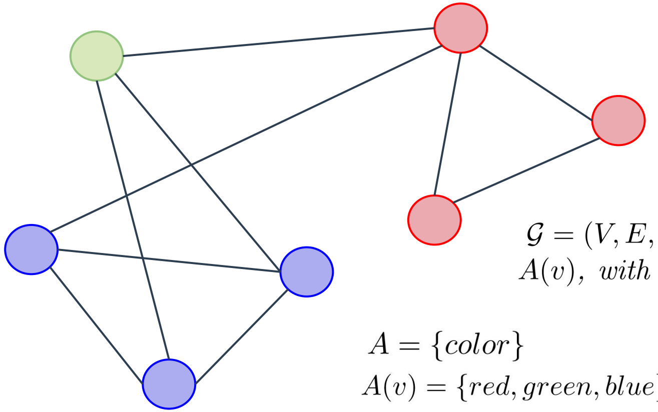
Salvatore Citraro

# A brief context (of my work)

Feature-rich networks (*Interdonato et al.* 2019)

→ More information as a complement to the topology

e.g. node-attributed networks



$$\mathcal{G} = (V, E, A)$$
$$A(v),\ with\ v \in V$$

$$A = \{color\}$$
$$A(v) = \{red, green, blue\}$$

Improve solutions to complex network tasks

→ **Community detection**: EVA [1]

**Network measures:** Conformity

[1] Citraro S., Rossetti G. (2020) "Eva: Attribute-Aware Network Segmentation". COMPLEX NETWORKS 2019

# Node-attributed networks

## What can we do?

Community detection

→ well-connectedness and **homogeneity**

Network measures

→ quantify **homophily** according to the attributes carried by the nodes

# Homophily

Tendency of similar nodes to interact with similar others

→ *social networks*: education, age, gender, work, etc.

*co-citation networks*: topics

*linguistic networks*: psycholinguistic variables of words

*Idea 1:*

*nodes with similar characteristics (degree, labels) are connected with a higher probability than expected*

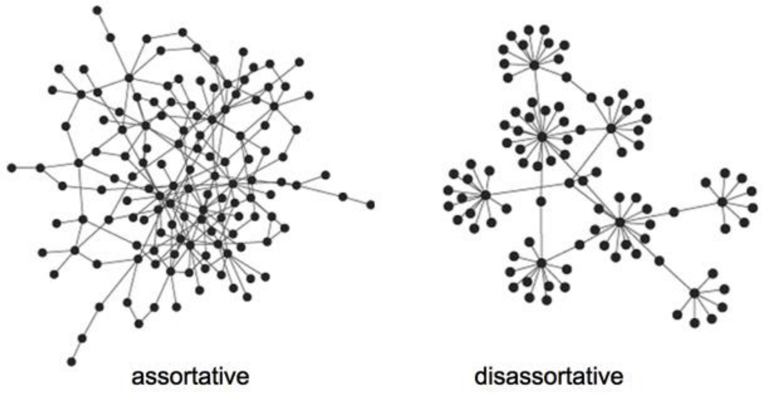
*Idea 2:*

*similar characteristics are more prominent along short distances*

# A special case: degree

## Newman's assortativity [2]

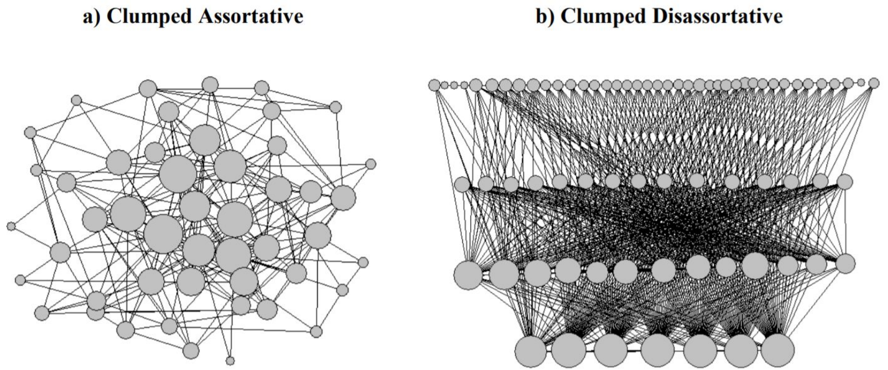
$$R = \frac{\sum_{xy} xy(e_{xy} - a_x \cdot b_y)}{\sigma_a \sigma_b}$$




[2] Newman, M. E. J. "Mixing Patterns in Networks." Physical Review E 67.2 (2003): n. pag. Crossref. Web.

## Clumpiness [3]

$$\Lambda(G,k,\alpha) = \sum_{i>j}^{n(n-1)/2} \frac{k_i k_j}{(d_{ij})^{\alpha}},$$




[3] Estrada, N. Hatano, A. Gutierrez, "Clumpiness mixing in complex networks", Journal of Statistical Mechanics: Theory and Experiment.

# Newman's assortativity (categorical)

A **global** measure based on Pearson's *r*

$$r = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i}$$

- *r* = -1 perfectly disassortative
- *r* = 0 no assortative (or random) mixing
- *r* = 1 perfectly assortative

| | | women | | | | |
|---|---|---|---|---|---|---|
| | | black | hispanic | white | other | $a_i$ |
| men | black | 0.258 | 0.016 | 0.035 | 0.013 | 0.323 |
| | hispanic | 0.012 | 0.157 | 0.058 | 0.019 | 0.247 |
| | white | 0.013 | 0.023 | 0.306 | 0.035 | 0.377 |
| | other | 0.005 | 0.007 | 0.024 | 0.016 | 0.053 |
| | $b_i$ | 0.289 | 0.204 | 0.423 | 0.084 | |

*r* = 0.621

***Limitation***

- An average quantification of mixing pattern across the *whole* network
- *Different patterns* and *outliers* are not identified

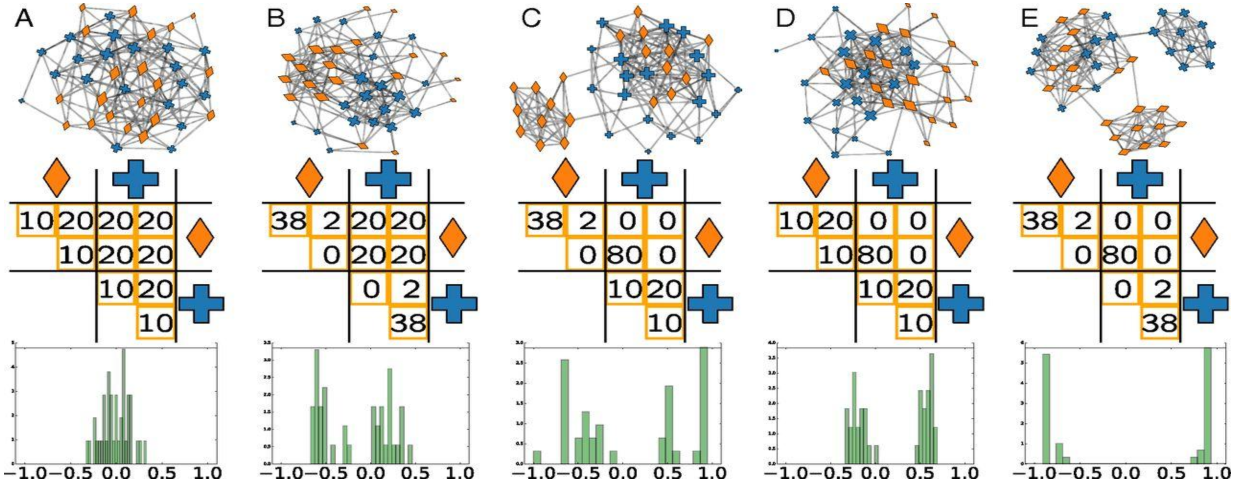
# Peel's assortativity [4]

A **node-centric** measure based on a multiscale strategy

→ overcome limits of global assortativity

$$e_{gh}(\ell) = \sum_{i:y_i=g} \sum_{j:y_j=h} w(i;\ell) \frac{A_{ij}}{k_i},$$

$$w_{\text{multi}}(i;\ell) = \int_0^1 w_\alpha(i;\ell)\,\mathrm{d}\alpha$$



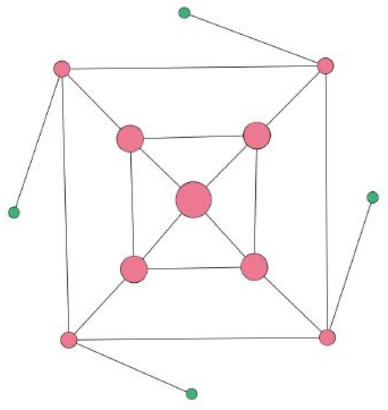

[4] L. Peel, J.-C. Delvenne, R. Lambiotte, "Multiscale mixing patterns in networks", Proceedings of the National Academy of Sciences.
Detailed explanation of the measure: https://piratepeel.github.io/slides/MixingPatterns_IC2S2.pdf

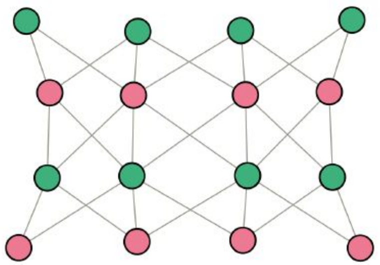
# Conformity (Rossetti G., Citraro S., Milli L.)

$$I_{u,v} = \begin{cases} 1 & \text{if } l_u = l_v \\ -1 & \text{otherwise} \end{cases}$$

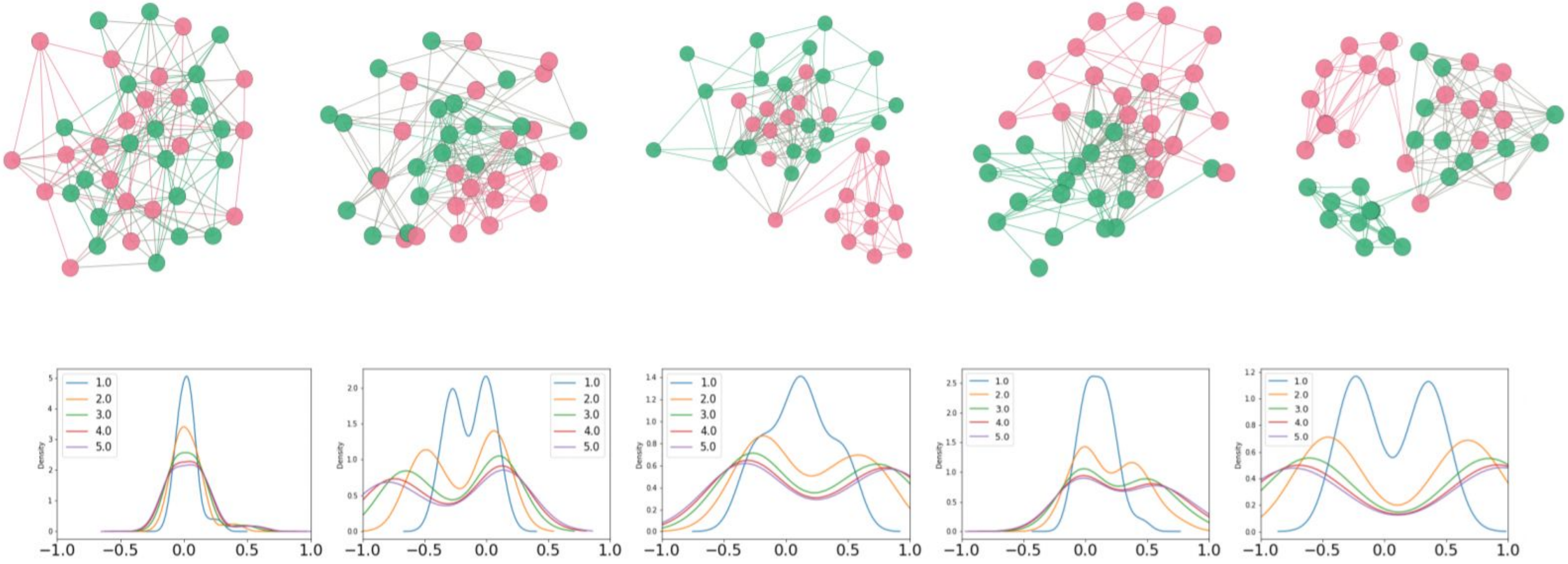
$$f_{u,l_u} = \begin{cases} \frac{|\{v | v \in \Gamma(u) \wedge l_u = l_v\}|}{|\Gamma(u)|} & \text{if } > 0 \\ 1 & \text{otherwise} \end{cases}$$

$$\psi(u,\alpha) = \frac{\sum_{d \in D} \frac{\sum_{v \in N_{u,d}} I_{u,v} f_{v,l_v}}{|N_{u,d}| d^{\alpha}}}{\sum_{d \in D} d^{-\alpha}}$$





# Conformity (cont'd)



# Case studies

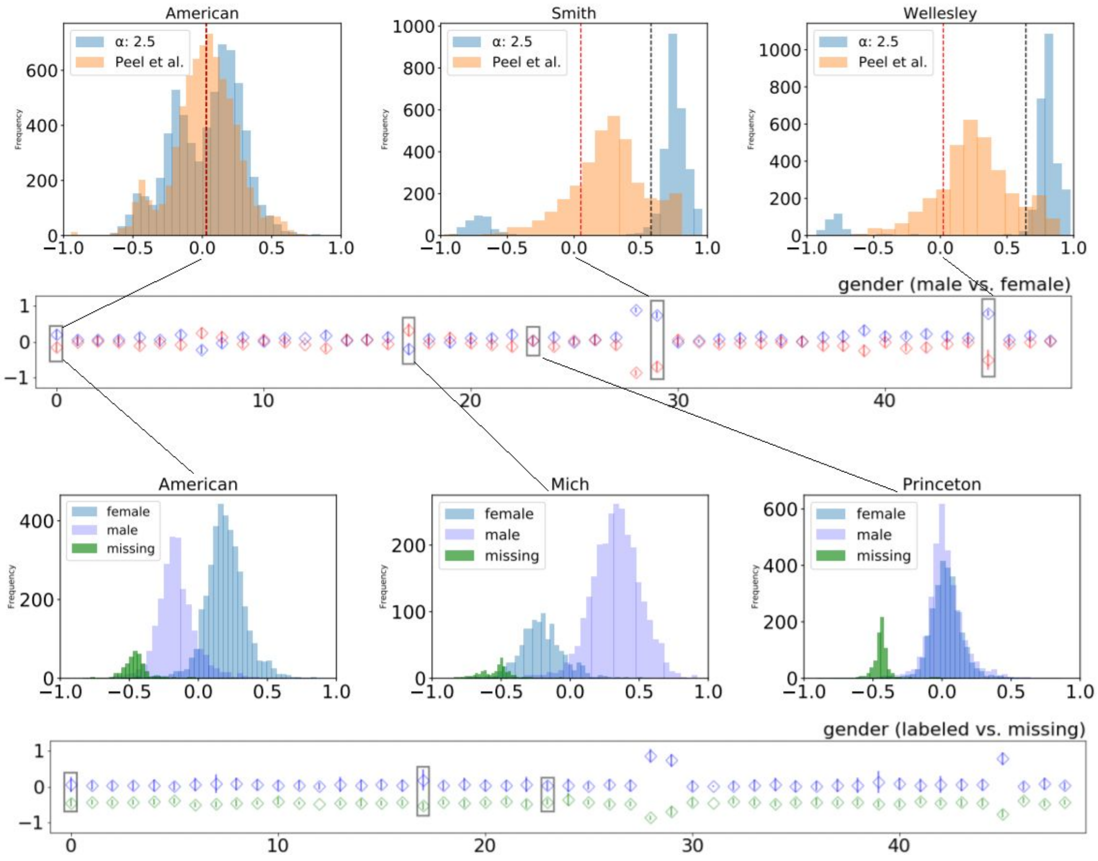
**Facebook100**: gender, year, dorm, etc...

→ just an overview of Conformity

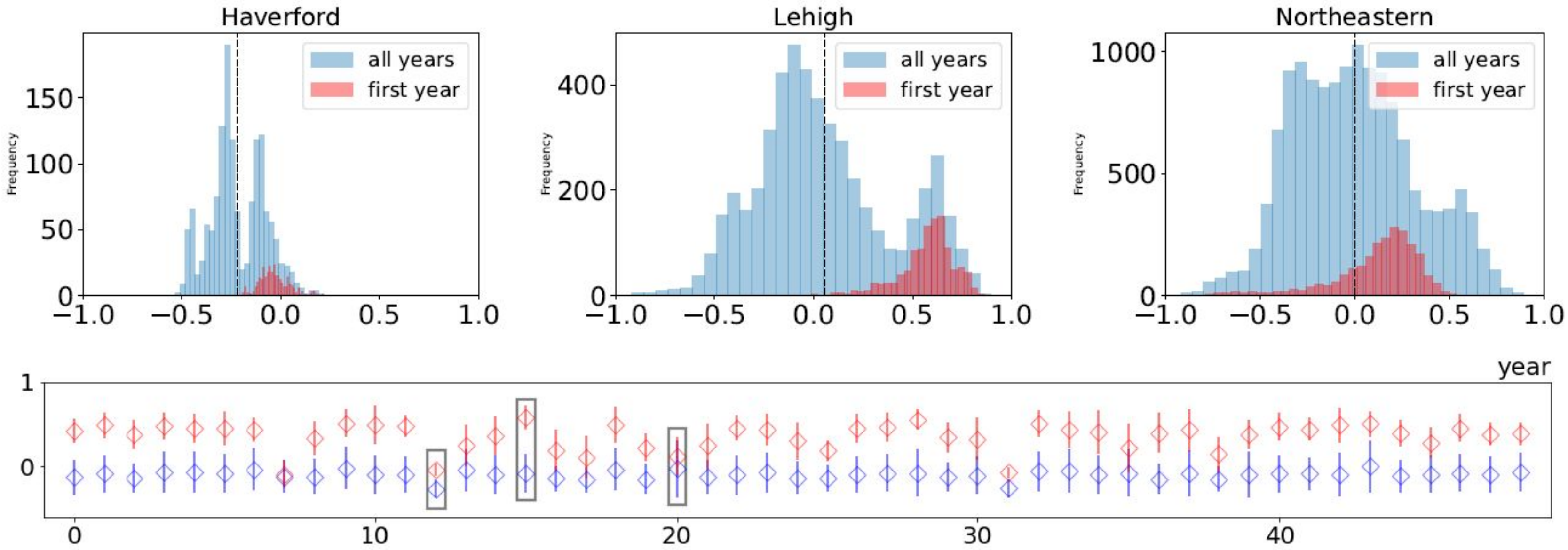
**Interaction data from Copenhagen Network Study**: gender

→ a statistically significant comparison of Conformity and Peel's assortativity

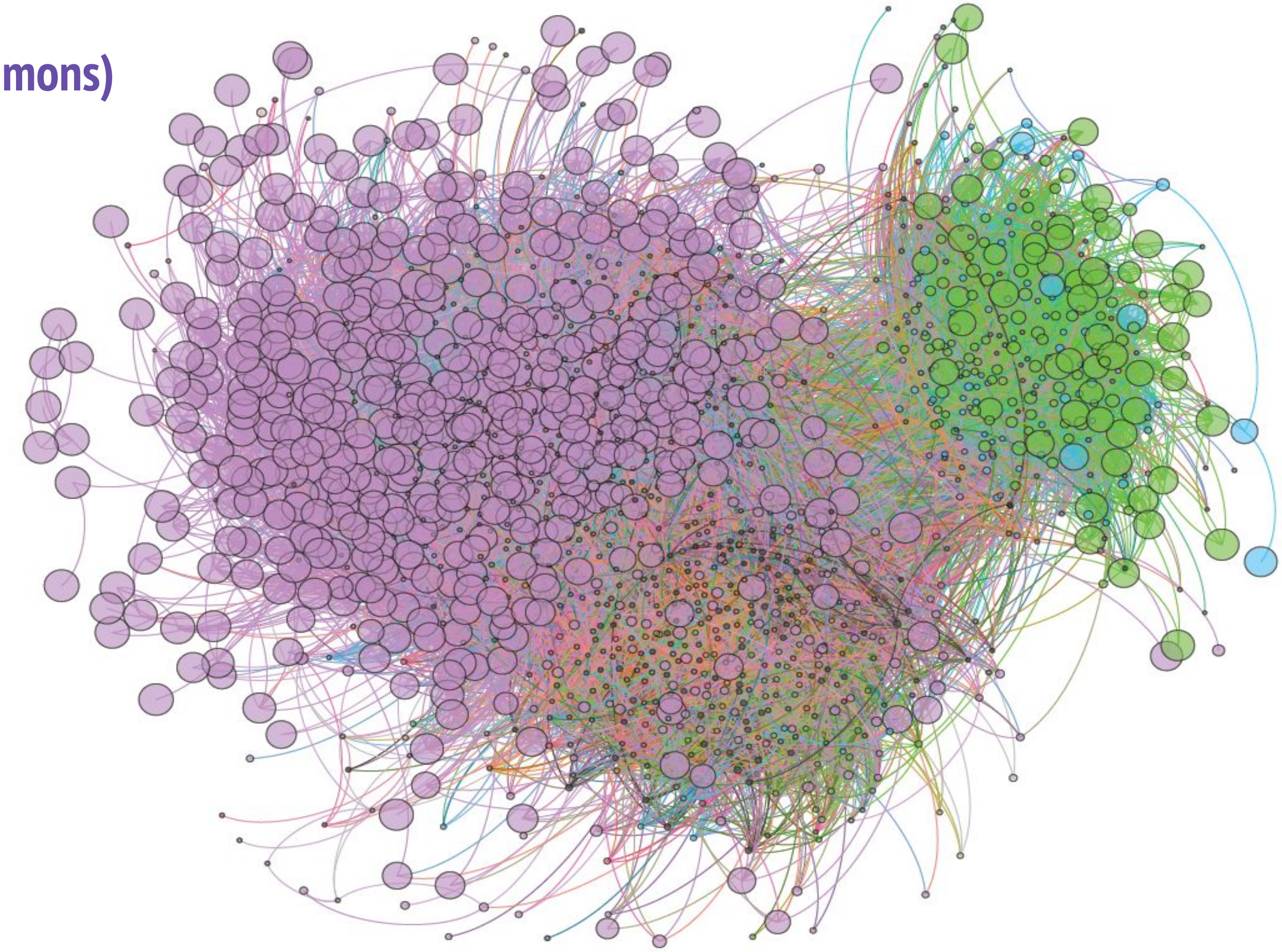
# Facebook100 - Gender



# Facebook100 - Year



# Dorm (Simmons)

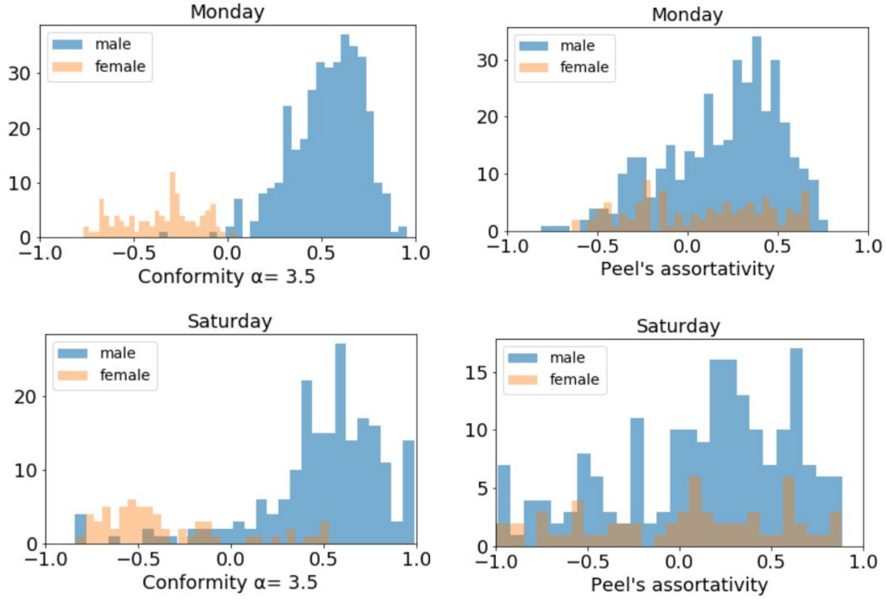


# Interaction data from Copenhagen Network Study

**Homophily by gender**:

→ the most difficult to capture

(under-representation of women, etc)




In the absence of a **ground truth**

we can not say whether Conformity

or Peel's assortativity approximate

the network behaviour

# Framework of comparison

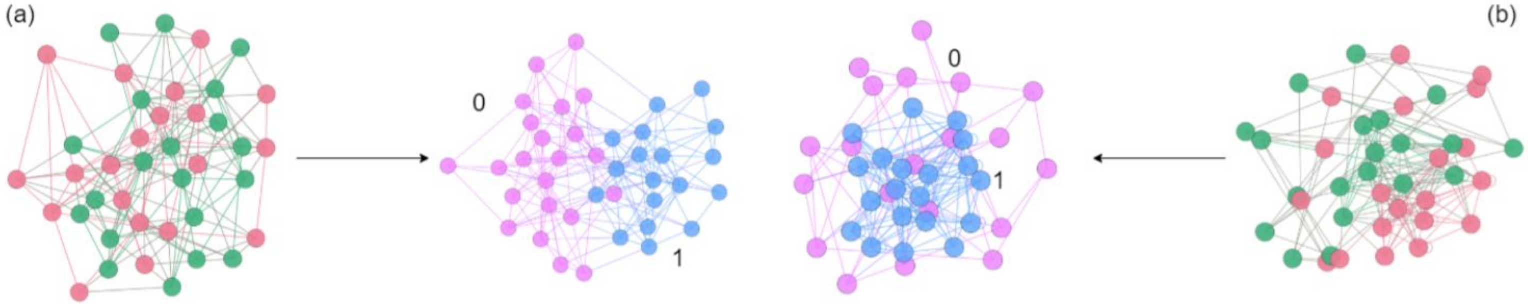
**Community structure** as a matter of comparison

→ the **minority group** within a community must be more heterophilic than the **majority group**

**Hypothesis**

→ the more gender groups are **unbalanced** within a community, the more the minority group is heterophilic w.r.t. gender

# Peel's quintet

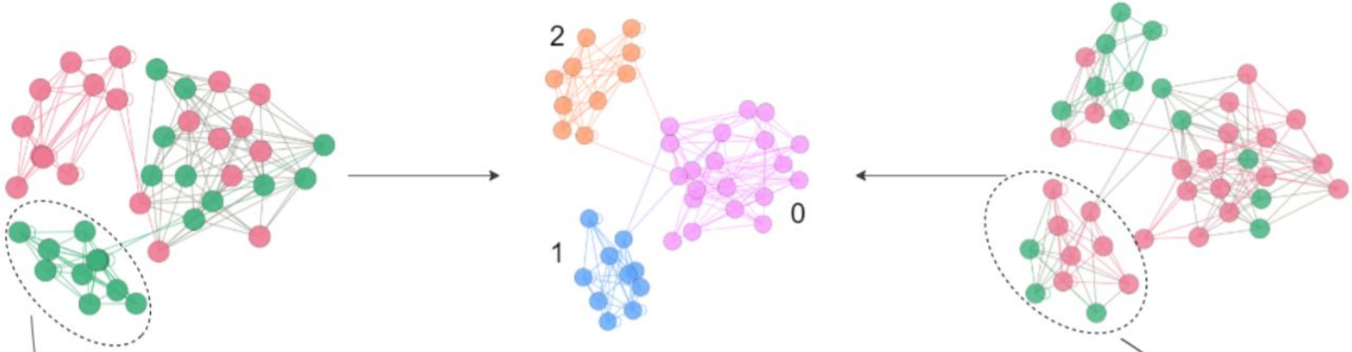



| com | label | N | Confrm. | Peel's ass. |
|---|---|---|---|---|
| 0 | g | 10 | 0.02 | 0.00 |
| | r | 12 | 0.11 | -0.01 |
| 1 | g | 10 | 0.03 | -0.04 |
| | r | 8 | -0.02 | -0.06 |

| com | label | N | Confrm. | Peel's ass. |
|---|---|---|---|---|
| 0 | g | 10 | -0.68 | -0.82 |
| | r | 10 | -0.71 | -0.81 |
| 1 | g | 10 | 0.12 | 0.25 |
| | r | 10 | 0.13 | 0.25 |

# Peel’s quintet (unbalanced)

(e)




| com | label | N | Confrm. | Peel’s ass. |
|---|---|---|---|---|
| 0 | g | 10 | -0.69 | -0.86 |
| | r | 10 | -0.68 | -0.85 |
| 1 | g | 10 | 0.91 | 0.89 |
| 2 | r | 10 | 0.91 | 0.89 |

| com | label | N | Confrm. | Peel’s ass. |
|---|---|---|---|---|
| 0 | g | 5 | **-0.44** | **-0.63** |
| | r | 15 | 0.12 | 0.05 |
| 1 | g | 3 | **-0.43** | **-0.23** |
| | r | 7 | 0.35 | 0.23 |
| 2 | g | 7 | 0.33 | 0.26 |
| | r | 3 | **-0.23** | **-0.10** |

# First step: find the core

A meta-definition of **community** is not enough

→ the comparison must be done with nodes strongly embedded within their communities

(statistically significant) **degree embeddedness**

$$e_i = k^{in} - k^{out}$$

$$t\_score(e_i) = \frac{e_i - \widehat{e}}{\frac{\sigma}{\sqrt{n}}}$$

# CNS - Monday (Walktrap)

First level

| | | Conformity | | Peel's assrt. | |
|---|---|---|---|---|---|
| | N | M | SD | M | SD |
| **Community** | | | | | |
| 0 | 130 | **0.333** | 0.442 | **0.130** | 0.339 |
| 1 | 92 | **0.036** | 0.252 | **0.112** | 0.318 |
| 2 | 75 | **0.298** | 0.397 | **0.158** | 0.316 |
| 3 | 66 | **0.551** | 0.320 | **0.251** | 0.291 |
| 4 | 46 | **0.542** | 0.392 | **0.201** | 0.322 |
| 5 | 17 | **0.475** | 0.327 | **0.190** | 0.438 |

t-score ($e_i$) > 0

| | | Conformity | | Peel's assrt. | |
|---|---|---|---|---|---|
| | N | M | SD | M | SD |
| **Core Community** | | | | | |
| 0 | 73 | **0.394** | 0.431 | **0.174** | 0.351 |
| 1 | 78 | **0.032** | 0.232 | **0.113** | 0.305 |
| 2 | 60 | **0.343** | 0.392 | **0.163** | 0.326 |
| 3 | 56 | **0.561** | 0.341 | **0.272** | 0.288 |
| 4 | 31 | **0.565** | 0.422 | **0.159** | 0.344 |
| 5 | 13 | **0.600** | 0.092 | **0.131** | 0.409 |

Second level

| | | | Conformity | | Peel's assrt. | |
|---|---|---|---|---|---|---|
| | | N | M | SD | M | SD |
| **Community** | **Gender** | | | | | |
| 0 | 0 | 103 | **0.546** | 0.147 | **0.145** | 0.326 |
| | 1 | 27 | **-0.475** | 0.172 | **0.075** | 0.386 |
| 1 | 0 | 50 | **0.234** | 0.143 | **0.114** | 0.289 |
| | 1 | 42 | **-0.200** | 0.107 | **0.109** | 0.363 |
| 2 | 0 | 59 | **0.486** | 0.172 | **0.196** | 0.303 |
| | 1 | 16 | **-0.390** | 0.143 | **0.016** | 0.330 |

| | | | Conformity | | Peel's assrt. | |
|---|---|---|---|---|---|---|
| | | N | M | SD | M | SD |
| **Core Community** | **Gender** | | | | | |
| 0 | 0 | 61 | **0.570** | 0.159 | **0.192** | 0.347 |
| | 1 | 12 | **-0.497** | 0.205 | **0.084** | 0.371 |
| 1 | 0 | 41 | **0.228** | 0.112 | **0.119** | 0.365 |
| | 1 | 37 | **-0.184** | 0.099 | **0.106** | 0.348 |
| 2 | 0 | 49 | **0.512** | 0.156 | **0.191** | 0.312 |
| | 1 | 11 | **-0.410** | 0.177 | **0.037** | 0.372 |

# Second step: analysis of variance

**Hard** due to group size unbalance itself

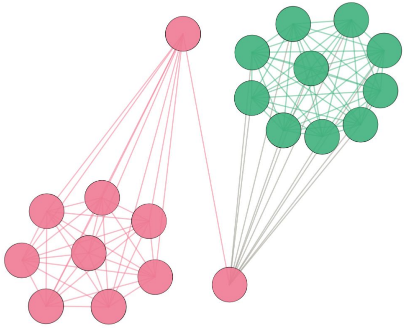
→ two-way ANOVA?

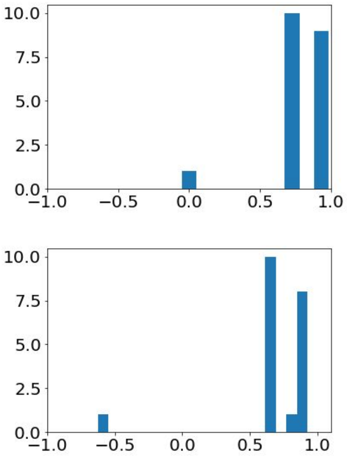
An idea: **Mann-Whitney U**

| | Conformity | | Peel's assortativity | |
|---|---|---|---|---|
| | Mann Whitney U | p-value | Mann Whitney U | p-value |
| Community 0 | 0.0 | 0.00 | 1234.0 | 0.10 |
| Community 1 | 38.0 | 0.00 | 1047.0 | 0.40 |
| Community 2 | 0.0 | 0.00 | 320.0 | 0.02 |

| | Conformity | | Peel's assortativity | |
|---|---|---|---|---|
| | Mann Whitney U | p-value | Mann Whitney U | p-value |
| Core community 0 | 0.0 | 0.00 | 294.0 | 0.14 |
| Core community 1 | 0.0 | 0.00 | 747.0 | 0.45 |
| Core community 2 | 0.0 | 0.00 | 204.0 | 0.10 |

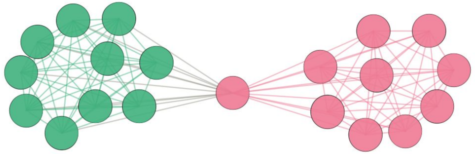
# Why?

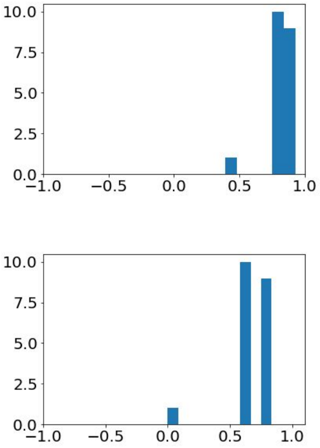





Peel’s assortativity

Conformity






Maybe *Peel’s assortativity* can not *scale* in extremely unbalanced situations

# Conclusion and future works

1. **Conformity** is more coherent than **Peel's assortativity** w.r.t. the community structure of networks (must be proven better in future)
2. Conformity is quite expensive
3. Is Conformity a **metrics**?