Structural Uniqueness in Networks

2021 Expert Meeting on Statistical Data Confidentiality

Mark van der Loo, Rachel de Jong, Frank Takes, Marieke de Vries, Peter-Paul de Wolf
Statistics Netherlands, Leiden University

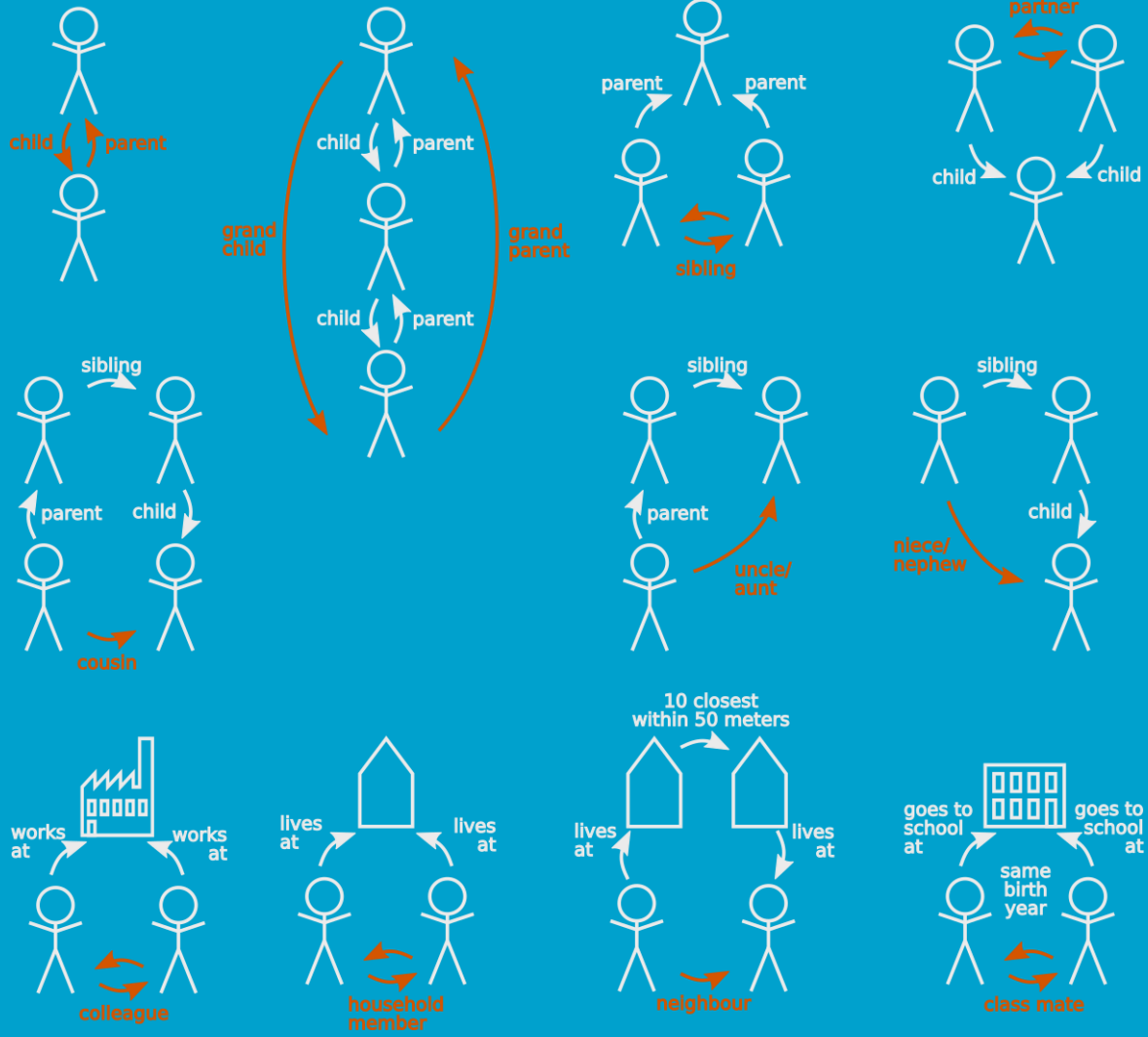
December 3rd 2021, Pòznan

Why networks?

NSIs have data available that make constructing networks possible

Connections between data entries can be valuable

Possibility to extract new information from dataset compared to traditional microdata



16.9 mln nodes

39 billion links

avg 1800 links pp

(van der Laan, de Jonge, 2017)

Risks of sharing network data

Risk measures for traditional microdata files are well known (e.g., *k*-anonymity)

Networks can furthermore contain:

- Personal information
- Relationships
- Personal information about relationships
- Relationships of relationships
- … etc

How risky?

Network structure is an additional attribute

Can a structure be identifying?

How do we translate *k*-anonymity to networks?

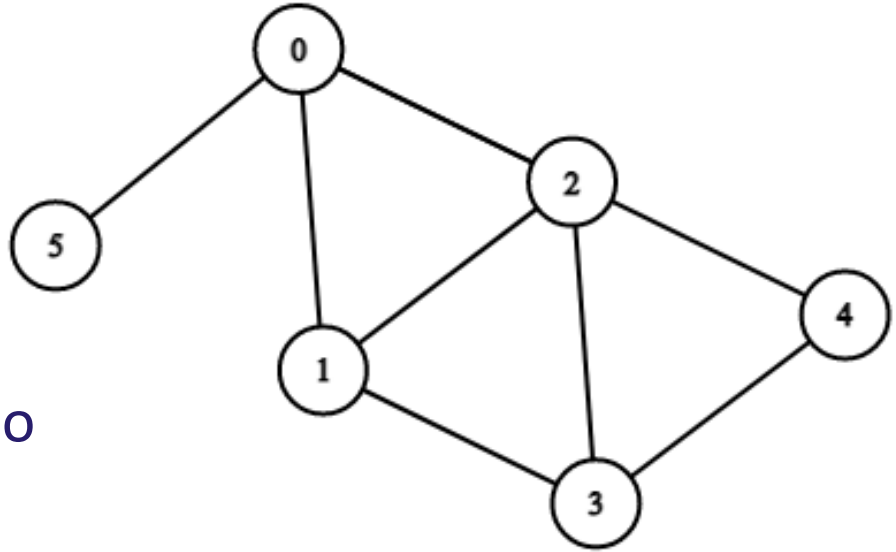
- *K*-automorphism (Zou et al., 2009): occuring in *k*-sized orbit of graph automorphism group. Takes into account complete structural position.
- Measure based on degree of node and surrounding nodes (Hay et al., 2008): Existence of cycles as revealing feature not taken into account.

What is the structure?

Unlabelled graph

Are any of the nodes unique enough?

How much information do we compare?



D-k anonymity

We measure structural uniqueness by comparing nodes with similar structural positions in the network

Node is *d-k* anonymous if there are *k* nodes with “identical structural position in their surrounding up to distance *d*”.

We call those ‘identical’ nodes *d*-equivalent.

D-equivalent

We define similar nodes by using isomorphisms

Two nodes v and w are d-equivalent when

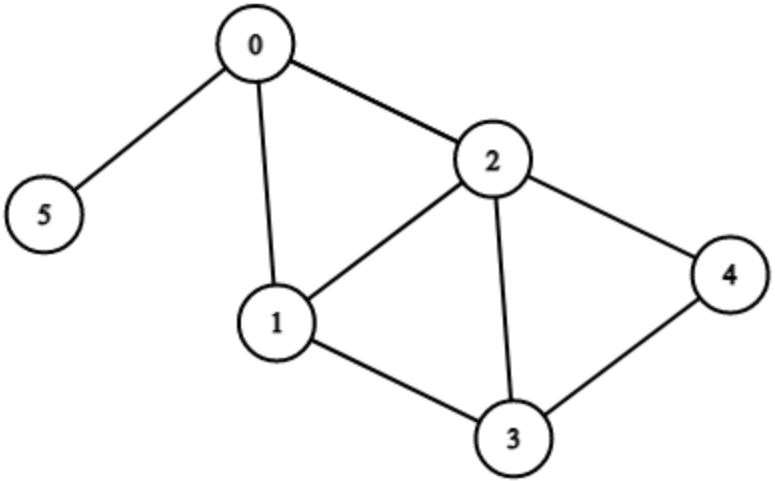
1. $N(v,d) \simeq N(w,d)$; and
2. There is an isomorphism $\phi: N(v,d) \rightarrow N(w,d)$ such that $\phi(v) = w$.

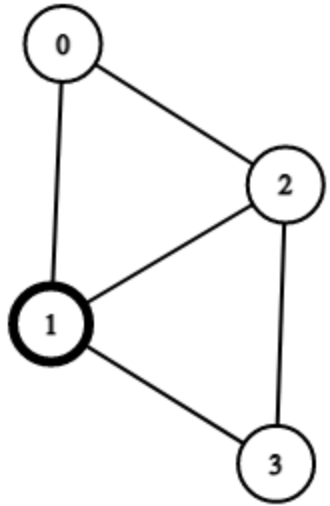
Where $N(v,d)$ is the neighbourhood of v up to distance d

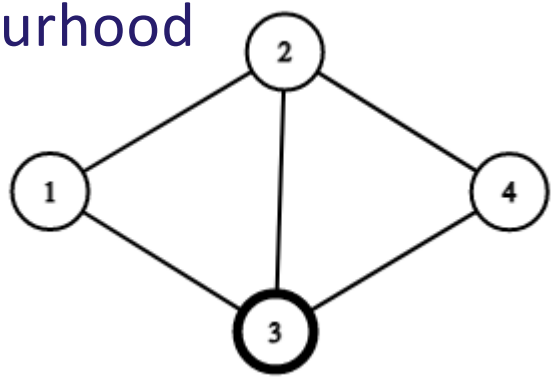
D-equivalent

Nodes are d-equivalent when they have the same structural position in their (d^{th} order) neighbourhood

Node 1 and 3 are 1-equivalent:



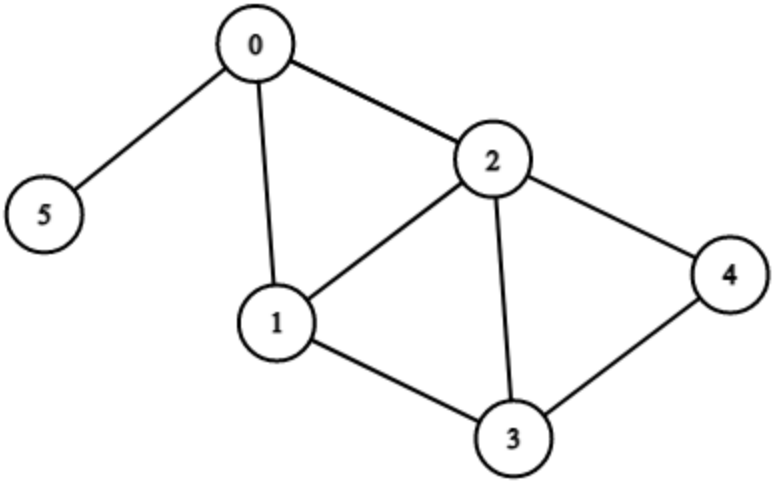


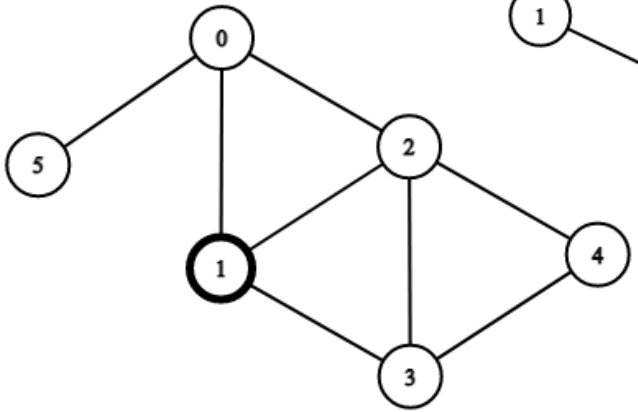


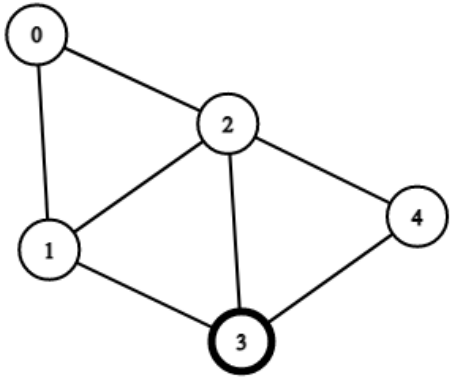
D-equivalent

Nodes are *d*-equivalent when they have the same structural position in their (d^{th} order) neighbourhood

Node 1 and 3 are **not** 2-equivalent:





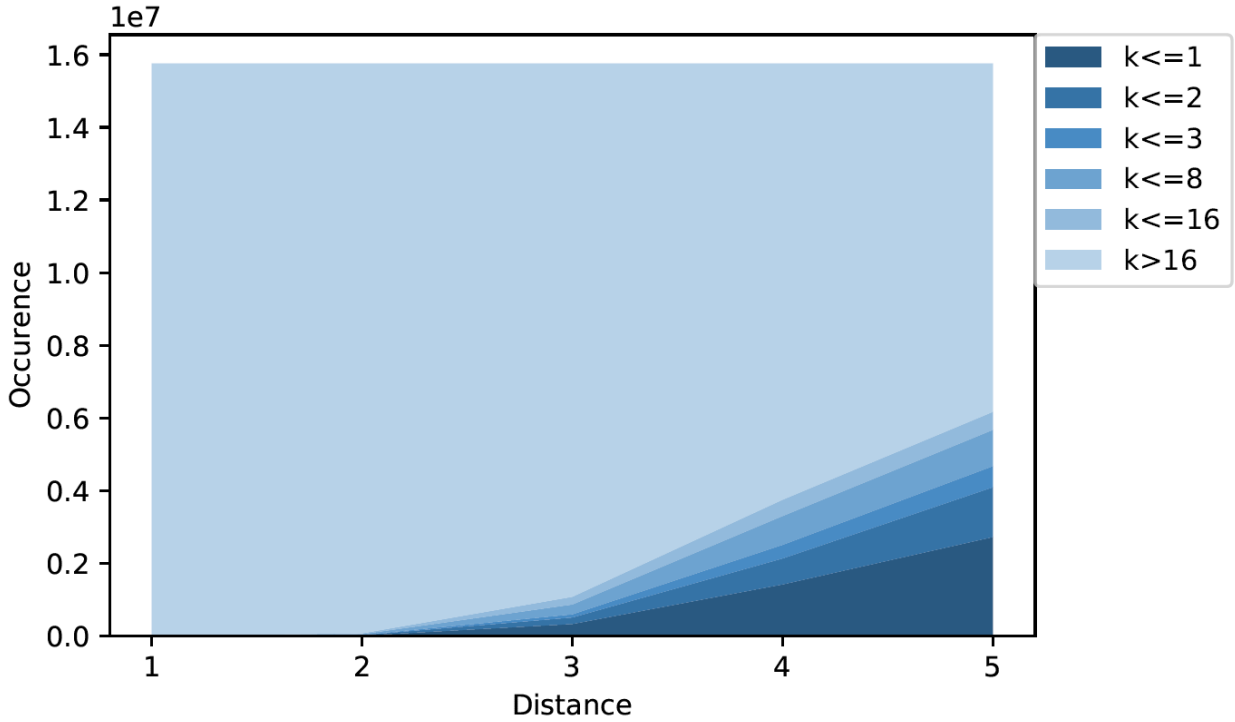


D-k anonymity

Node *v* is *d-k* anonymous if for *d* there are *k*-1 nodes for which:

1. The d^{th} order neighbourhood is isomorphic to that of *v* (i.e. same number and ordering of nodes)
2. There is an isomorphism that maps the node to *v* (i.e. the node plays the same role in the neighbourhood)

Results on family network



Reproduced from De Jong (2021)

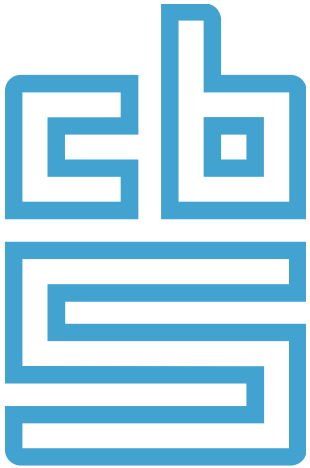
To conclude

- In order to protect and share networks, we need to be able to measure the associated risk
- We have defined a measure able to check the structural uniqueness of nodes
- Uniqueness of nodes grows quickly by the size of the neighbourhood

Current research

Interested in:

- Adding labels to nodes
- Adding “fuzzy” matching



Facts that matter

References

- van der Laan, J., M. Das, S. te Riele, E. de Jonge, and T. Emery (2021). *Measuring educational segregation using a whole population network of the netherlands*. https://osf.io/preprints/socarxiv/7jtb2/.
- Hay, M., G. Miklau, D. Jensen, D. Towsley, and P. Weis (2008). *Resisting structural re-identication in anonymized social networks.* Proceedings of the VLDB Endowment 1 (1), 102-114.
- Zou, L., L. Chen, and M. T. Özsu (2009). *K-automorphism: A general framework for privacy preserving network publication.* Proceedings of the VLDB Endowment 2 (1), 946-957.
- De Jong, R. G. (2021). *Measuring structural anonymity in complex networks.* Master's thesis, Leiden University, The Netherlands.