Analysis of the Southern Women Network using Fractional Approach

Vladimir Batagelj
IMFM Ljubljana, IAM UP Koper, and NRU HSE Moscow

1293. Sredin seminar
on Zoom, February 17, 2021

Outline

Vladimir Batagelj: vladimir.batagelj@fmf.uni-lj.si
Current version of slides (February 17, 2021 at 03 : 57): slides PDF
https://github.com/bavla/Linked/blob/master/docs

Davis' Southern women network

The Southern women network was collected in 1930s by Davis et al. (1941). It is one among the most analyzed networks in the SNA literature. Freeman (2003) published a meta-analysis in which he reviewed 21 partitions of the set of women obtained by different approches. In this paper we present another another analysis based on fractional approach.

The Southern women network is a two-mode network $\mathcal{N} = ((P, E), L)$ describing the social activities of 18 women (first mode, set of persons P) whom they observed over a nine-month period. During that period, various subsets of these women had met in a series of 14 informal social events (second mode, set of events E) (UCI, 2008). The participation of women in events determines the set of links L

$$(p, e) \in L \Leftrightarrow \text{woman } p \text{ attended event } e.$$

Two versions of network

Names of Participants of Group I	Code Numbers and Dates of Social Events Reported in *Old City Herald*													
	(1) 6/27	(2) 3/2	(3) 4/12	(4) 9/26	(5) 2/25	(6) 5/19	(7) 3/15	(8) 9/16	(9) 4/8	(10) 6/10	(11) 2/23	(12) 4/7	(13) 11/21	(14) 8/3
1. Mrs. Evelyn Jefferson	×	×	×	×	×	×	....	×	×	....	....	....	....	....
2. Miss Laura Mandeville	×	×	×	....	×	×	×	×	....	....	....	....	....	....
3. Miss Theresa Anderson	....	×	×	×	×	×	×	×	×	....	....	....	....	....
4. Miss Brenda Rogers	×	....	×	×	×	×	×	×	....	....	....	....	....	....
5. Miss Charlotte McDowd	....	....	×	×	×	....	×	....	....	....	....	....	....	....
6. Miss Frances Anderson	....	....	×	....	×	×	....	×	....	....	....	....	....	....
7. Miss Eleanor Nye	....	....	....	....	×	×	×	×	....	....	....	....	....	....
8. Miss Pearl Oglethorpe	....	....	....	....	....	×	....	×	×	....	....	....	....	....
9. Miss Ruth DeSand	....	....	....	....	×	....	×	×	×	....	....	....	....	....
10. Miss Verne Sanderson	....	....	....	....	....	....	×	×	×	....	....	×	....	....
11. Miss Myra Liddell	....	....	....	....	....	....	....	×	×	×	....	×	....	....
12. Miss Katherine Rogers	....	....	....	....	....	....	....	×	×	×	....	×	×	×
13. Mrs. Sylvia Avondale	....	....	....	....	....	....	×	×	×	×	....	×	×	×
14. Mrs. Nora Fayette	....	....	....	....	....	×	×	....	×	×	×	×	×	×
15. Mrs. Helen Lloyd	....	....	....	....	....	....	×	×	....	×	×	×	....	....
16. Mrs. Dorothy Murchison	....	....	....	....	....	....	....	×	×	....	....	....	....	....
17. Mrs. Olivia Carleton	....	....	....	....	....	....	....	....	×	....	×	....	....	....
18. Mrs. Flora Price	....	....	....	....	....	....	....	....	×	....	×	....	....	....

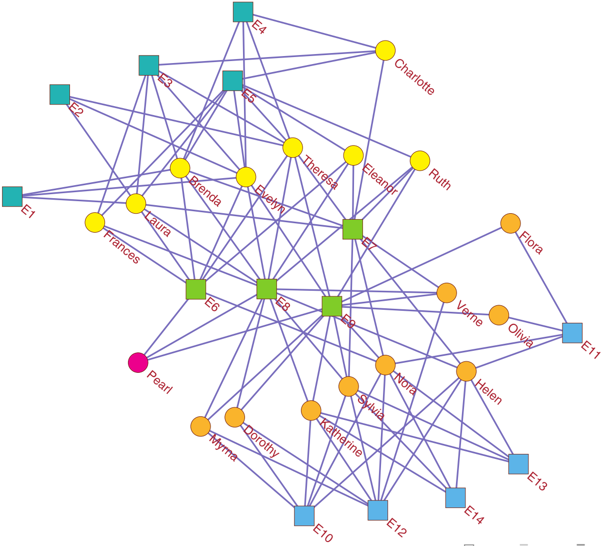
Type of Membership	Members	Events and Participations													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
Clique I:															
Core	1	C	C	C	C	C	C	–	C	C					
	2	C	C	C	–	C	C	C	C	–					
	3	–	C	C	C	C	C	C	C	C					
	4	C	–	C	C	C	C	C	C	–					
Primary	5			P	P	P	–	P	–	–					
	6			P	–	P	P	–	P	–					
	7					P	P	P	P	–					
Secondary	8					–	S	–	S	S					
Clique II:															
Secondary	9					S	–	S	S	S					
	10							S	S	S	–	–	S		
Primary	11							–	P	P	P	–	P		
	12							–	P	P	P	–	P	P	P
Core	13							C	C	C	C	–	C	C	C
	14						C	C	–	C	C	C	C	C	C
	15							C	C	–	C	C	C	C	C
Secondary	16								S	S	S	–	S		
	17									S	–	S			
	18									S	–	S			

In the book Davis et al. (1941) two versions of the Southern women network are given. We use the second version (also included in Ucinet). In the first version four links (Dorothy, E10), (Dorothy, E12), (Helen, E13), and (Helen, E14) are missing. Also, in the original data, Myrna is called Myra.

The matrix representation (the ordering of rows/persons and columns/events) provides an early attempt of blockmodeling of two-mode networks.

Southern Women – graph

spring embedder



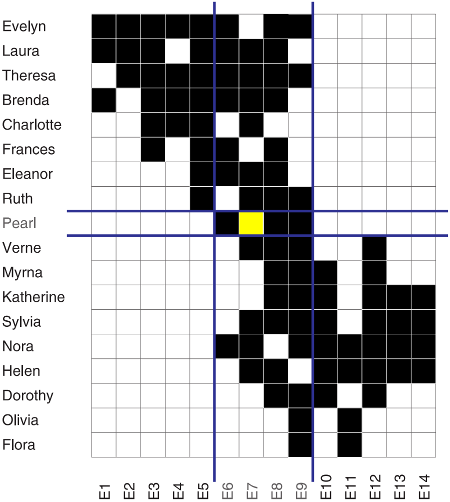
Clusterings from graph

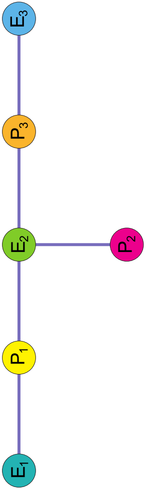
$$
\begin{aligned}
P_1 &= \{\text{Frances, Laura, Brenda, Evelyn, Theresa, Eleanor, Ruth, Charlotte}\},\\
P_2 &= \{\text{Pearl}\},\\
P_3 &= \{\text{ Myrna, Dorothy, Katherine, Sylvia, Nora, Helen, Verne, Olivia, Flora}\},\\
E_1 &= \{\text{E1, E2, E3, E4, E5}\},\\
E_2 &= \{\text{E6, E7, E8, E9}\},\\
E_3 &= \{\text{E10, E11, E12, E13, E14}\}.
\end{aligned}
$$

If we join the clusters P_1 and P_2 into a single cluster $P_1 \cup P_2$ we get a regular blockmodel with an empty row (error) for Pearl in the block $(P_1 \cup P_2) \times E_1$. All other blocks are regular without errors.

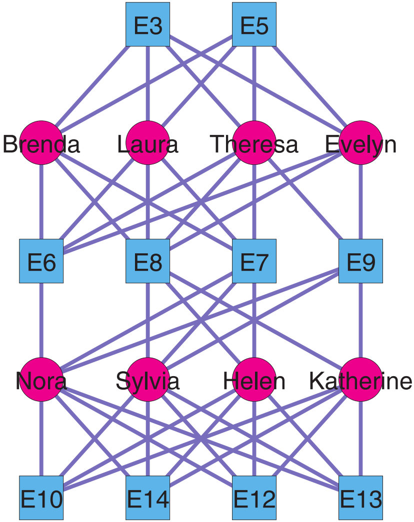
For the first version of the network, the partition gives a regular blockmodel with an additional empty row (error) for Dorothy in the block $P_3 \times E_3$.

Regular blockmodel





Two-mode cores



The Southern women network contains a (5,4)-core (Cerinšek and Batagelj, 2015) on node subsets $P_{5,4} = \{$ Brenda, Laura, Theresa, Evelyn, Nora, Sylvia, Helen, Katherine $\}$ and $E_{5,4} = \{$ E3, E5, E6, E7, E8, E9, E10, E12, E13, E14 $\}$. Inside the core each woman attended at least 5 events and each event was attended by at least 4 women.

The (4,6)-core is induced by the node subsets $P_{4,6} = \{$ Brenda, Laura, Eleanor, Theresa, Evelyn, Ruth, Nora, Sylvia, Helen, Myrna, Verne, Dorothy, Katherine $\}$ and $E_{4,6} = \{$ E5, E6, E7, E8, E9, E10, E12 $\}$.

Projections

A traditional (classic) approach (Breiger, 1974; Borgatti and Halgin, 2014) to analysis of a binary two-mode network $\mathcal{N}$ represented with its matrix $\mathbf{PE} = [PE[p,e]]_{P \times E}$ is its transformation to the corresponding one-mode networks:

- *row projection* (projection to the row set P):
  $\mathbf{PP} = \text{row}(\mathbf{PE}) = \mathbf{PE} \cdot \mathbf{PE}^T$
- *column projection* (projection to the column set E):
  $\mathbf{EE} = \text{col}(\mathbf{PE}) = \mathbf{PE}^T \cdot \mathbf{PE}$

These networks are further analyzed using standard methods.

Properties of projections

The entry $PP[p, q]$ counts the number of ways of moving from person p to person q through some event from E – the number of co-attended events

$$PP[p, q] = |PE(p) \cap PE(q)|$$

where $PE(p) = \{e \in E : PE[e, p] \neq 0\}$ is the set of neighbors in the set E of the node $p \in P$. It holds $\text{outdeg}(p) = |PE(p)|$.

Similarly, $EE[e, f] = |PE(e) \cap PE(f)|$,
$PE(e) = \{p \in P : PE[e, p] \neq 0\}$ and $\text{indeg}(e) = |PE(e)|$.

For the standard projections it turns out that each node from the other set contributes a complete subgraph on the neighbors in the projection set – nodes with large degree are over-represented in the projection (Batagelj, 2020).

Normalized projection

To make the contribution of each node equal, in bibliometrics proposed the fractional approach (Price and Beaver, 1966) based on the *normalized matrix* $n(\mathbf{PE}) = [n(PE)[p,e]]$ where

$$n(PE)[p,e] = \frac{PE[p,e]}{\max(1, \mathrm{indeg}(e))}$$

The *normalized projection* $\mathbf{PP}_n$ to the set P is computed as

$$\mathbf{PP}_n = n(\mathbf{PE}) \cdot n(\mathbf{PE})^T$$

The projection matrix is symmetric, $PP_n[p,q] = PP_n[q,p]$. The corresponding network $\mathcal{N}_P = (P, L_P, w_P)$ is undirected with edges (undirected links) $(p:q) = (q:p)$

$$(p:q) \in L_P \Leftrightarrow PP_n[p,q] > 0$$

The weight of the edge $(p:q)$ is set to $w_P((p:q)) = 2PP_n[p,q]$ for $p \neq q$; and $w_P((p:p)) = PP_n[p,p]$. It measures the fractional co-attendance contribution of women p and q.

Newman's normalization

In the case that we would like to have a "fair" projection without loops we can adapt the normalized projection using the *Newman's normalization* $n'(\mathbf{PE}) = [n'(PE)[p,e]]$ where

$$n'(PE)[p,e] = \frac{PE[p,e]}{\max(1, \mathrm{indeg}(e) - 1)}$$

The *loopless normalized projection* $\mathbf{PP}_N$ to the set P is computed as

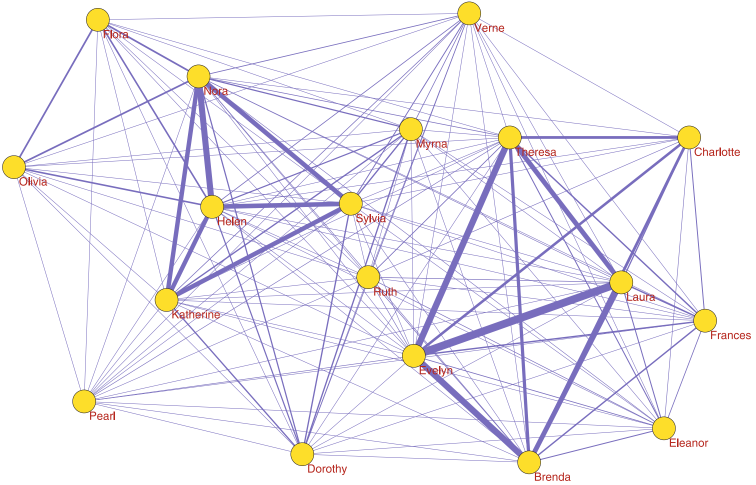
$$\mathbf{PP}_N = n(\mathbf{PE}) \cdot n'(\mathbf{PE})^T$$

The projection matrix is symmetric, $PP_N[p,q] = PP_N[q,p]$. The corresponding network $\mathcal{N}'_P = (P, L'_P, w'_P)$ is undirected with edges (undirected links) $(p:q) = (q:p)$

$$(p:q) \in L'_P \Leftrightarrow p \neq q \wedge PP_N[p,q] > 0$$

We remove loops. The weight of the edge $(p:q)$ is set to $w'_P((p:q)) = 2PP_N[p,q]$.

Loopless normalized projection to women



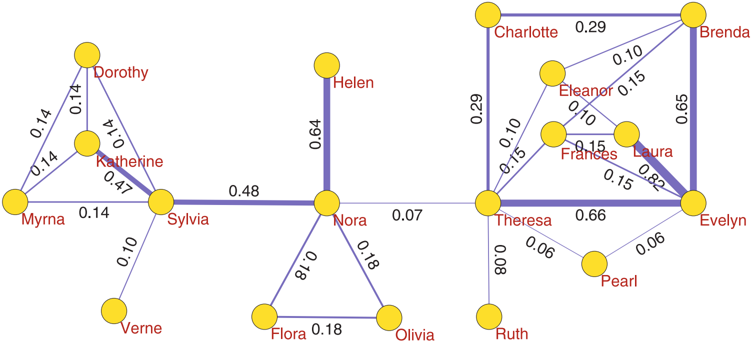
Pathfinder skeleton

The picture of the obtained loopless normalized projection network is not easy readable. To simplify it we determine its *pathfinder skeleton* for weights $w_{PF} = 1/w_P$ (the PF weight must be a dissimilarity) and $r = \infty$ (see Schvaneveldt et al. (1988) and Batagelj et al. (2014), p. 96–102) preserving the connectivity and the most important edges.

For a given weight d that measures a dissimilarity the Pathfinder algorithm of order $r = \infty$ removes from the undirected network $\mathcal{N}$ all edges $(x : y)$ from triangles on nodes x, y, z with edges satisfying the condition $\max(d(x,z), d(z,y)) < d(x,y)$. The edges of the skeleton on the following slide have original weights w_P.

```
Network/Create New Network/Transform/Line Values/Power [-1]
Network/Create New Network/Transform/Reduction/Pathfinder* [0]
Select PF network as Second
Select original network as First
Networks/Cross-Intersection/First
```

Women – Pathfinder skeleton



From the skeleton we obtain by cuting the weakest edge (Nora : Theresa) a partition in two clusters:

$P_1 = \{$Myrna, Dorothy, Verne, Katherine, Sylvia, Nora, Helen, Olivia, Flora$\}$,
$P_2 = \{$Theresa, Evelyn, Brenda, Charlotte, Laura, Eleanor, Frances, Pearl, Ruth$\}$.

Partition of events

Mathematically we could either compute the matrix $n(\mathbf{PE})^T \cdot n(\mathbf{PE})$ or (duality) repeat the same procedure on the matrix $\mathbf{EP} = \mathbf{PE}^T$ and get the normalized projection to the set of events E. Unfortunately, it doesn't work well. The problem is the meaningfulness (interpretation) of the corresponding projection. To get a kind of normalized projection to the set of events E we apply the approach proposed in Batagelj (2020).

Let's compute the matrices $\mathbf{Pn} = \mathbf{PE} \cdot n(\mathbf{PE})^T$ and $\mathbf{nP} = n(\mathbf{PE}) \cdot \mathbf{PE}^T$. The entry

$$Pn[p,q] = \sum_{e \in PE(p) \cap PE(q)} PE[p,e] \cdot n(PE[q,e])$$

measures the fractional co-attendance contribution of woman q to joint events with woman p; and the entry $nP[p,q]$ measures the fractional co-attendance contribution of woman p to joint events with woman q.

Jaccard column projection

A *joint co-attendance* contribution can be defined as an average of these two values. Among different options we selected one that gives the *Jaccard column projection* $Jcol(\mathbf{PE})$

$$Jcol[e,f] = \frac{|PE(e) \cap PE(f)|}{|PE(e) \cup PE(f)|}$$

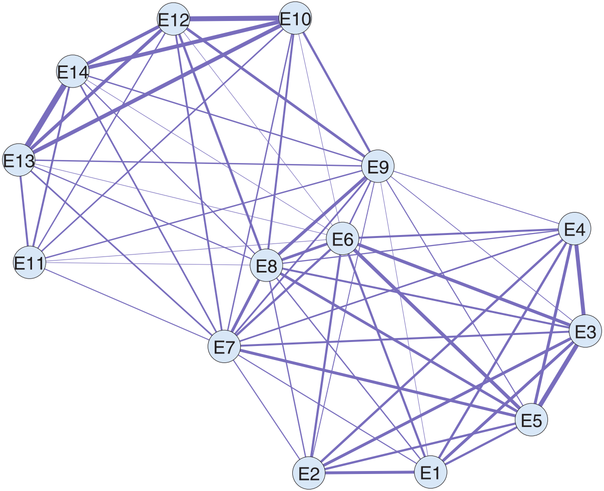
$Jcol$ is symmetric, $Jcol[e,f] = Jcol[f,e]$, and $Jcol[e,e] = 1$.

Considering $|PE(e) \cap PE(f)| = EE[e,f]$ and $|PE(e) \cap PE(f)| + |PE(e) \cup PE(f)| = |PE(e)| + |PE(f)| = \mathrm{indeg}(e) + \mathrm{indeg}(f)$ we get

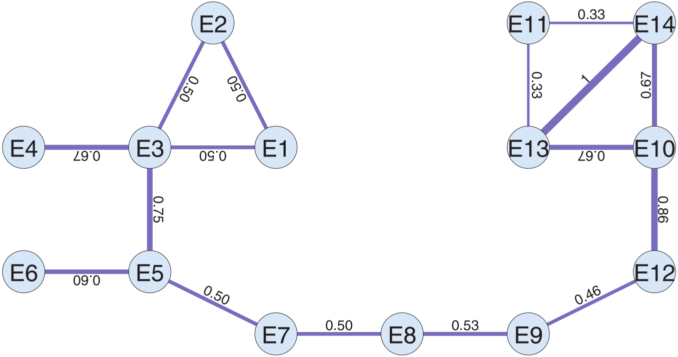
$$Jcol[e,f] = \frac{EE[e,f]}{\mathrm{indeg}(e) + \mathrm{indeg}(f) - EE[e,f]}$$

The Jaccard column projection can be obtained from the ordinary column projection. The corresponding Jaccard network $\mathcal{J} = (E, L_J, Jcol)$ is undirected with loops removed.

Jaccard network



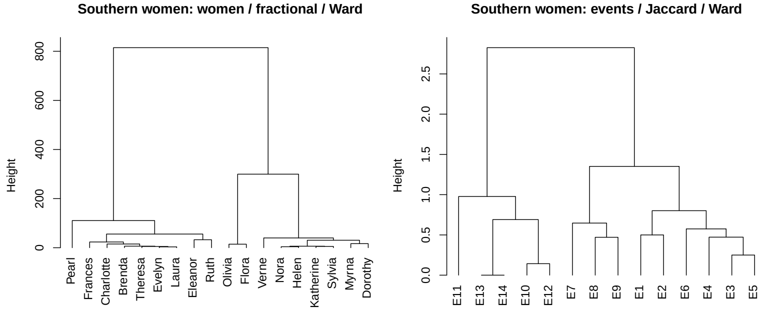
Events – Pathfinder



Applying again the Pathfinder algorithm to $\mathcal{J}$ (*Jcol* is a similarity measure, $w_{PF} = 1 - Jcol$) we get a skeleton. Removing two weakest edges (E9, E12) and (E5, E7) we get three clusters of events

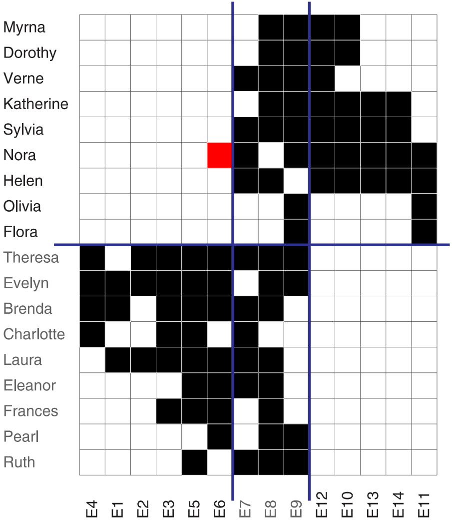
$E_1 = \{$ E4, E1, E2, E3, E5, E6 $\}$,
$E_2 = \{$ E7, E8, E9 $\}$,
$E_3 = \{$ E12, E10, E13, E14, E11 $\}$.

Hierarchical clusterings



The partitions of the sets P and E can be obtained also using the hierarchical clustering of the corresponding dissimilarities. From the dendrograms we get the same partitions as from the Pathfinder skeletons.

Blockmodel


Myrna
Dorothy
Verne
Katherine
Sylvia
Nora
Helen
Olivia
Flora
Theresa
Evelyn
Brenda
Charlotte
Laura
Eleanor
Frances
Pearl
Ruth
E4
E1
E2
E3
E5
E6
E7
E8
E9
E12
E10
E13
E14
E11


E3
P1
E2
P2
E1

Blockmodel

The obtained regular blockmodel is very simple:

- each woman attended at least one event from the cluster E_2;
- each woman from the cluster P_2 also attended at least one event from the cluster E_1, and no event from the cluster E_3 (except Nora);
- each woman from the cluster P_1 also attended at least one event from the cluster E_3, and no event from the cluster E_1.

We also searched for regular (prespecified) two-mode blockmodels of the Southeren women network using the optimizational approach (Doreian et al., 2004). In most cases (number of clusters) we obtained several exact (no error) solutions but without an interesting interpretation.

Conclusions

- In the paper we proposed a new approach to blockmodeling of affiliation / two-mode networks (Persons, Events) based on the fractional approach from bibliometrics. We illustrated it by analyzing the Southern women network. We defined the fractional similarity measure between Persons and the Jaccard similarity measure between Events, both based on fractional co-attendance. To find the clusterings of Persons / Events we can use either the Pathfinder algorithm or the hierarchical clustering. The obtained clusterings determine the blockmodel.
- The Freeman's review of the Southern women network analyses deals only with the partitions of the set of women. We considered equally also the partitioning of the set of events and obtained a simple but meaningful regular two-mode blockmodel.

Acknowledgments

All computations were performed using the program for large network analysis and visualization Pajek (De Nooy et al., 2018) and the statistical programming system R.

This work is supported in part by the Slovenian Research Agency (research program P1-0294 and research projects J1-9187 and J5-2557), and prepared within the framework of the HSE University Basic Research Program.

References I

Batagelj, V. (2020): On fractional approach to analysis of linked networks. Scientometrics 123 (2): 621-633

Batagelj, V., Doreian P., V., Ferligoj, A., Kejžar N. (2014): Understanding Large Temporal Networks and Spatial Networks: Exploration, Pattern Searching, Visualization and Network Evolution. Wiley.

Borgatti, S.P., Halgin, D.S. (2014): Analyzing Affiliation Networks. Chapter 28 in The SAGE Handbook of Social Network Analysis. John Scott, Peter J. Carrington (Eds.), Sage.

Breiger, R. L. (1974): The duality of persons and groups. Social Forces, Vol. 24: 201-229.

Cerinšek, M., Batagelj, V. (2015): Generalized two-mode cores. Social Networks 42, 80–87.

Davis, A., Gardner, B. B. and Gardner, M. R. (1941) Deep South. Chicago: The University of Chicago Press.

De Nooy, W., Mrvar, A., Batagelj, V. (2018): Exploratory Social Network Analysis with Pajek; Revised and Expanded Edition for Updated Software. Structural Analysis in the Social Sciences, CUP.

Doreian, P, Batagelj, V, Ferligoj, A (2004): Generalized blockmodeling of two-mode network data. Social Networks 26 (1): 29–53.

References II

Egghe, L., Rousseau, R. (1990): Introduction to informetrics, quantitative methods in library, Documentation and information science. Amsterdam: Elsevier

Freeman, L.C. (2003): Finding social groups: A meta-analysis of the southern women data, In Breiger, R., Carley, K. and Pattison, P. eds. Dynamic Social Network Modeling and Analysis. Washington: The National Academies Press.

Price, D. J. de Solla, Beaver, D. de (1966): Collaboration in an invisible college. American Psychologist, 21(11), 1011-1018.

Schvaneveldt, R. W., Dearholt, D. W., Durso, F. T. (1988): Graph theoretic foundations of Pathfinder networks. Comput. Math. Applic. 15 (4), 337-345

Southern Women Data Set. UCIrvine Network Data Repository, 2008. https://networkdata.ics.uci.edu/netdata/html/davis.html