Social Stratification from Networks of Leveling Ties

David Schoch[1] and Ulrik Brandes[2]

[1] Department of Sociology, The University of Manchester, United Kingdom,
david.schoch@manchester.ac.uk
[2] Chair of Social Networks, ETH Zurich, Switzerland,
ulrik.brandes@gess.ethz.ch

Abstract. Social networks can be made of various kinds of ties, but (often implicit) assumptions embodied in network-analytic tools do not necessarily apply to all of them. Centrality indices, for instance, build on the assumption that it is always beneficial to add more ties. While it has been noted that networks of ties with a negative sentiment require different concepts of centrality, we here highlight ties that are neither positive nor negative to have, but an indication of commonality. This is exemplified by the derivation of socio-economic status from networks that indicate common class membership.

Keywords: social stratification, centrality, leveling ties, interval graphs

1 Introduction

As Wasserman and Faust note, "[...] the range and type of ties can be quite extensive." [39]. Indeed, an edge or tie in a network can have a multitude of meanings, ranging from friendship relations to complex protein–protein interactions in cells [2]. Additionally, any of these ties can have a direction, carry a weight or even connect more than two nodes.

The overarching principle of ties in most empirical networks, however, is that they have a positive connotation. That is, accumulating ties is always beneficial and forming new connections is better than not to. Many of the existing network analytic tools were developed under this premise. As a prime example, consider the concept of centrality. Most axiomatic approaches for centrality are designed by assuming that adding an edge can never make a node less central [33]. More recently, it was shown that the neighborhood-inclusion preorder underlies many index-induced rankings [34]. Across the board, being connected to the same, and potentially more, nodes than some other node, renders the former node more central.

A conceptually different and less frequently studied class are signed networks, consisting of a mixture of positive and negative ties. The analytical focus in such networks is mostly distinct from studies with exclusively positive ties. A reoccurring research question for such networks is related to *structural balance theory* [21]. A signed network is defined to by structurally balanced if every cycle has an even number of negative ties [3].

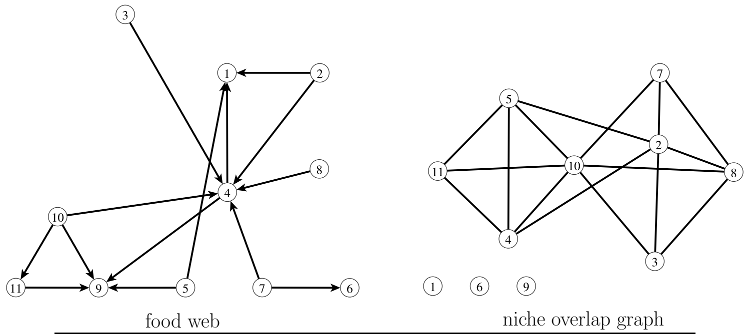
In this work, we consider a third general type of a tie, which is neither positive nor negative. We refer to these ties as *leveling ties*, reflecting the fact that they indicate a form of equivalence, equality or, in a weaker sense, similarity between nodes. We argue, that existing tools designed for networks with positive or negative ties are not applicable to these kind of networks and require conceptually different approaches to assess typical network related questions such as centrality.
The main goal of this work is to introduce the concept of leveling ties and outline first steps toward analyzing such networks. While we heavily draw from existing methods in the literature, they have thus far not been considered in conjunction and only in one dimensional settings.
We start by introducing leveling tie by means of some example networks and idealized network structures in Section 2. In Sections 3 and 4, we compile a first set of tools to analyze such networks, where the main focus is set on obtaining node orderings. Section 5 is devoted to an application in the context of social stratification, which constitutes the main empirical contribution of this work. We conclude with a brief discussion in Section 6.

2 Networks of leveling ties

In this section, we review existing networks from the literature with leveling ties and discuss idealized network structures. The idealized structures later guide the choice and design of methods to analyze such networks coherently.

2.1 Empirical networks

In the context of ecology, networks with leveling ties occur, for instance, in so-called *niche overlap graphs* [5]. Given a *food web*, that is a network which summarizes predator-prey relations of an ecosystem, two species are connected in the niche overlap graph if and only if they have common prey. The underlying assumption is that if two species have a common prey, then their respective ecological niches overlap. That is, some of the environmental factors that ensure the species survival coincide. An example of such a graph is shown in Figure 1.
A second example of a network with leveling ties is shown in Figure 2. The network represents co-mentioning of authors in any published literary review in the Swedish press during 1881-1893 [31]. Pairs of authors that are linked in the network can be taken to occupy a common literary niche, similar to the niche overlap in ecological context.
While the two examples are taken from very different domains, they do share some commonalities. Nodes in the network are embedded in an n-dimensional latent space, and each node is represented in each dimension by a range of values. A species in an ecosystem, for instance, can be characterized by the range of tolerable environmental factors that ensure its survival (e.g. a limited range of temperatures, light and moisture). Thus, each species is identified by a box, or niche, in the latent n-dimensional space. The same holds for authors, who can be characterized in such



1 Canopy	7 Middle-zone scansorial animals
2 Canopy animals	8 Middle-zone flying animals
3 Upper air animals	9 Ground
4 Insects	10 Small ground animals
5 Large ground animals	11 Fungi
6 Trunk, fruit, flowers	

Fig. 1. Food web and niche overlap graph of the Malaysian Rain Forest, reprinted from [30].

a space based on literary factors (e.g. aesthetics, religious, moral and political aspects).

A tie between nodes occurs if the ranges overlap in one or more dimension. These ties are conceptually different from the usual encountered ties in network research. They do not posses any positive or negative features since the notion of "sharing living space" or "literary commonalities" do not necessarily imply a beneficial or antagonistic interaction. Hence, it is neither desirable to accumulate many such ties nor avoid their creation entirely.

2.2 Idealized networks

Leveling ties can occur in diverse settings, but follow similar formation processes. Nodes are represented as a box in a latent space and networks are constructed by identifying the overlaps of the respective boxes. If boxes overlap, we assume that there exists a form of equivalence between the nodes involved. The underlying space, though, may be arbitrarily complex. In fact, any graph with n vertices can be viewed as an *intersection graph* of boxes in an n-dimensional space [29]. The question is, if all n dimensions are necessary to unambiguously characterize the position of a node. The *boxicity* of a graph is the minimum number of dimensions in which it can be represented as an intersection graph [29]. Ideally, the boxicity of a network with leveling ties is small and, in the best case, only one dimension is required. Such graphs are known as *interval graphs*, since nodes can be represented as intervals on a single axis [9]. A simple example is shown in Figure 3.

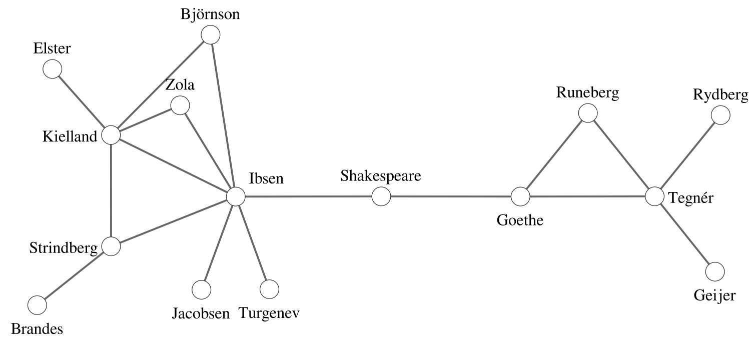


Fig. 2. Co-mentioning of authors in any published literary review in the Swedish press during 1881-1893 [31]. Authors are connected if they were co-mentioned more than five times and the proportion of co-mentions is more than three standard errors above the expectation.

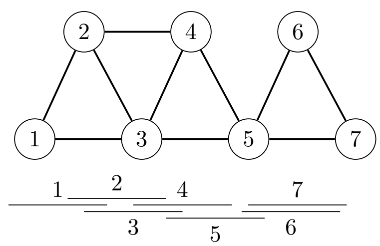


Fig. 3. Illustration of an interval graph and its one dimensional embedding.

Note that while each set of intervals has a unique representation as interval graph, the converse does not hold true [9]. Nevertheless, uncovering an underlying interval representation is the main goal for our setting, since it allows us to order nodes along a single dimension and compare their relative positioning.
Surprisingly, many empirically observed niche overlap graphs in ecology are interval graphs [5]. This finding greatly facilitates empirical research on ecosystems since only a single dimension needs to be investigated. Although this dimension is unspecified, it allows to order species along it and compare their relative positions.
Generally, we expect networks with leveling ties to deviate from being perfect interval graphs. Assessing the degree of this deviation is an important step prior to any analytic examination. The closer the network is to being an interval graph, the more structural properties we expect to coincide with those of an interval graph.

3 Measures of intervality

Interval graphs can be recognized in linear time [20]. In most use cases, however, it does not suffice to know if a graph is interval or not. If a

graph is not interval, it is desirable to have a method at hand which can assess the *intervality* of the graph. That is, how close is a graph to be interval? A potential measure can be derived via graph edit distances, i.e. how many edges must be added/deleted in order to turn a graph into an interval graph. The more edges need to be changed, the lower the intervality of a network. There exist at least three feasible instantiation of the interval edit problem: *interval graph completion* (only edge additions allowed), *interval graph deletion* (only edge deletions allowed), and *interval sandwich* (both edits allowed). All three are, however, NP-hard [15, 16, 18].
Apart from edit distances, there exist a variety of tractable heuristics stemming from the research on food webs and niche overlap graphs [4, 7, 37]. The following definitions are needed to define any such approach.

Definition 1. *A binary matrix A has the* consecutive ones property *(C1P) if there exists a permutation matrix Π such that the ones form a consecutive sequence in each column of ΠA.*

Definition 2. *The* maximal clique-vertex incidence matrix *M of an undirected graph $G = (V, E)$ is a $k \times n$ matrix, where k is the number of maximal cliques and n the number of vertices. An entry m_{ij} is one if the clique i contains vertex j and zero otherwise.*

The following theorem relates the C1P and the maximal clique-vertex incidence matrix to the class of interval graphs.

Theorem 1 ([13]). *A graph G is an interval graph if and only if its maximal clique-vertex incidence matrix M has the C1P.*

If there does not exist a permutation matrix Π such that a binary matrix A has the C1P, then there exist columns where 0's are surrounded by 1's. These gaps are referred to as *Lazarus events*. The *Lazarus count* $\ell(A)$ is defined as the minimal number of Lazarus events over the set of all (row) permutations. More formal,

$$\ell(A) = \min_{\Pi \in \mathcal{S}_n} \ell(\Pi A) \tag{1}$$

where

$$\ell(\Pi A) = \left|\left\{(\pi(i), j) \ : \ a_{\pi(i)j} = 0 \land \exists \pi(k) < \pi(i) < \pi(l) \text{ with } a_{\pi(k)j} = a_{\pi(l)j} = 1\right\}\right|$$

and $\mathcal{S}_n$ is the symmetric group. Note that the Lazarus count can equivalently be defined for rows or simultaneous row and column permutations. Additionally, the minimizing permutation is not unique. If a permutation π minimizes the Lazarus count $\ell(A)$ of a matrix A, then so does the reverse of π.
The Lazarus count of the maximal clique-vertex incidence matrix M of a graph G can serve as a heuristic to assess the intervality of a network. The higher its count, the lower the intervality. It remains, however, to determine the permutation matrix Π which minimizes this count. The preferred approach for food webs is *simulated annealing* [7, 37]. The method is, however, non-deterministic and results may vary between consecutive

runs. Additionally, an enormous number of iterations is needed which makes the method computationally expensive.

A different approach is based on the Laplacian matrix of MM^T [1]. It was shown that if M has the C1P then the permutation induced by ordering the Fiedler vector of MM^T yields a permutation matrix $\Pi^{(L)}$ such that $\ell(\Pi^{(L)}M) = 0$. If M does not have the C1P, then $\Pi^{(L)}$ at least gives a good approximation of the true minimal Lazarus count.

Graphs with boxicity greater than one are studied much less as associated problems are intractable. Even testing for boxicity two is NP-complete [22]. To the best of our knowledge, there exists only one rather cumbersome method to identify graphs with boxicity two [28].

4 Centrality indices

Most existing centrality indices evaluate trajectories such as paths or walks within a network [10, 33]. For leveling ties, it is not clear how to interpret paths and walks: are two non-adjacent neighbors of a node to be considered equivalent by transitivity or non-equivalent by the absence of a tie? Additionally, it is questionable if the intuition "adding ties can not reduce a nodes centrality", formalized in axiomatic approaches, holds in this context [33, 34].

Instead of evaluating structural importance in the traditional interpretation of centrality, we are rather interested in obtaining an ordering of nodes along a small set of key dimensions, preferably one if the network structure permits it. This ordering should, as for interval graphs, reflect an embedding of actors into an underlying latent space, where the dimensions represent factors that need to be interpreted by a domain expert.

The Fiedler vector approach is one possibility to obtain such an ordering, as illustrated in the previous section. Note that larger values do not necessarily mean that a node is more central, since the reverse ordering is equally valid. The importance of the ordering is not the specific rank, but rather the relative positions along a spectrum.

The Fiedler vector alone, however, is insufficient if the network has a low intervality score and thus a higher dimensionality. In such cases, we can use the next larger eigenvectors of the Laplacian matrix to obtain orderings for the higher dimensions. The network shown in Figure 2, for instance, has boxicity two [12]. We can therefore include the second smallest non-zero eigenvector of the Laplacian matrix to obtain an ordering for the second dimension. The result is shown in Figure 4.

Including the second smallest eigenvector evidently allows to differentiate authors that are very close in the first dimension. While the exact meaning of both dimensions remains unclear, the result shows that the Fiedler vector alone can not adequately capture the underlying relative positions of nodes in this case. Hence, the second smallest eigenvector is needed to reveal additional information, which is sufficient if the network has boxicity two.

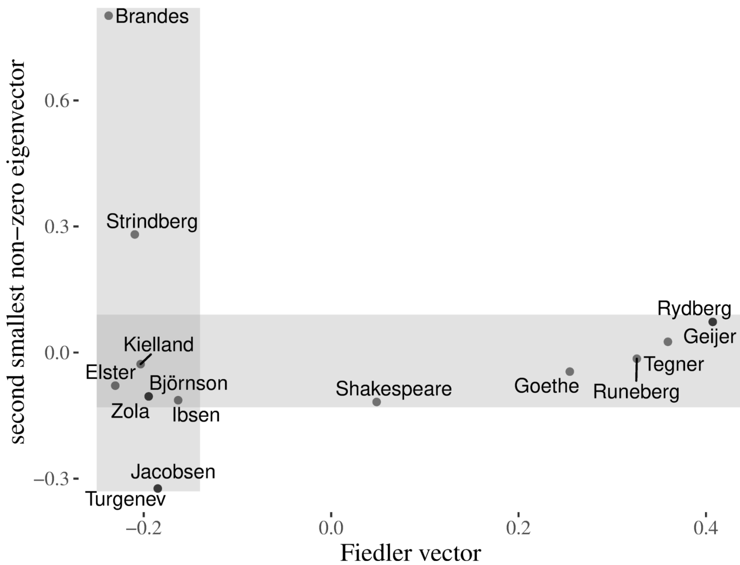


Fig. 4. Fiedler vector and second smallest non-zero Laplacian eigenvector of the Swedish literary network. Grey rectangles highlight the two-dimensional aspect of the ordering.

5 Application

In this section, we apply the previously introduced methods to the problem of social stratification. Our data basis constitutes occupational information of married or co-habitating couples derived from US census data.[3] We note that the potential of using network analytic tools to study social stratification was already recognized in [19]. However, the authors mainly employ it as an exploratory and complementary tool to existing methods.

5.1 Cambridge social interaction and stratification scale

A plethora of different measures exist to rank or classify occupations based on different socio-economic indicators. Prominent approaches include categorical measures which put occupations into a set of classes [8, 17], prestige rankings based on surveys [38] and socio-economic status scores [14]. More details can be found in reviews [6, 23].
A prominent measure which does not fit the previous categories is the *Cambridge social interaction and stratification scale* (CAMSIS), a representative for social distance approaches. It is based on the assumption that the presence of social classes can be inferred from how people group in everyday life [24, 25]. Relationships of individuals are embedded in socially moderated networks, wherein individuals interact in various ways.

[3] data obtained from https://usa.ipums.org/usa/ [32]

Relations outside of these networks are assumed to be substantively different. Friends and marriage partners, for instance, tend to be chosen more frequently from an individual's own social network [26]. CAMSIS scores were initially based on friendship ties as the central social interaction [35, 36]. More contemporary instantiations are derived from frequency tables of occupational combinations of married (or cohabitating) couples [27]. The underlying assumption is that frequently occurring combination of occupations should be located within close proximity in the social space. This frequency table is then projected into a two dimensional space using correspondence analysis.[4] Empirical findings suggest that the first dimension can be interpreted as an indicator of inequality or stratification among occupations. The higher the value for an occupation, the higher it is ranked in the stratification hierarchy.

5.2 Leveling ties between occupations

Incumbents of occupations are characterized by a range of socio-economic variables, such as level of education or income. Similar to species in ecology, occupations can thus be described as a box (or a "social niche") within a space of socio-economic factors. These social niches naturally overlap and thus offer opportunities for incumbents to interact socially, forming cordial and potentially also romantic relationship. We thus expect more marriage ties among people when their respective social niches overlap, much like it is highlighted by the social distance approach.

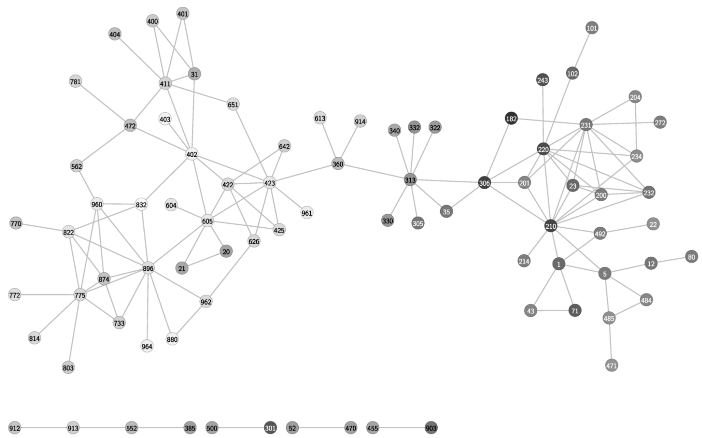
To determine leveling ties among occupations, we employ the following procedure. Two occupations are connected if the combination occurs at least in 1 out of 10,000 marriages and more than twice as likely as expected by chance given the size of the occupational groups. Figure 5 shows the resulting network derived from the US census data of 2000,[5] together with the CAMSIS scores of occupations.[6]

The Fiedler vector method yields an upper bound of 95 for the Lazarus count, indicating that the network is not representable in one dimension. To assess the significance of the observed count, we generated 5000 networks with the same degree sequence and computed the respective Lazarus counts. The Shapiro-Wilk test suggests that we can not reject the null-hypothesis of the Lazarus count to be normally distributed ($p > 0.5$). Additionally, the probability of observing a network with a smaller Lazarus count as the observed network is negligible ($p < 0.001$). We can thus conclude that the intervality of the observed network is significantly higher than expected. Using the method proposed in [28], we find that the network actually has boxicity two such that occupations can be ordered along two dimensions. Figure 6 shows the ordering based on the Fiedler vector and the next larger Laplacian eigenvector. The figure illustrates, that the Fiedler vector is able to discriminate between occupations with high and low CAMSIS scores. The second dimension allows for distinguishing occupations within these two groups.

[4] more advanced versions also employ *Goodman's Class of RC-II Association Models*

[5] occupational codes retrieved from https://usa.ipums.org/usa/volii/occ2000.shtml

[6] retrieved from http://www.camsis.stir.ac.uk/Data/USA.html



id	occupation	CAMSIS	id	occupation	CAMSIS
182	Psychologists	87.1	604	Graders & Sorters	30.2
306	Physicians & Surgeons	81.8	402	Cooks	31.0
210	Lawyers	80.9	423	Houskeeping Cleaners	31.4

Fig. 5. Network of occupations derived from US census data of 2000. Node color indicates the CAMSIS scores of occupations. The darker the node, the higher the score. The table shows the three occupations with the highest (left) and lowest (right) CAMSIS scores.

6 Discussion

Networks of leveling ties are surprisingly prevalent in the literature, yet have never explicitly been defined as such. These networks possess an intriguing connection to interval graphs which are among the best studied discrete structures in combinatorics. Freeman already pointed out in the 1980's that interval graphs have great potential in networks analysis, for instance when analyzing co-citation networks [11]. Indeed, carefully constructed co-citation networks can be seen as networks of leveling ties. Studying its dimensionality may reveal the complexity of a specific field and allow to differentiate among sub-domains by ordering authors along a small set of key dimensions. We actually expect the interpretation of leveling ties to apply to many networks that arise from a projection of a two-mode network. Affiliation networks are an obvious example, but more theoretical and empirical work needs to be done.

Assessing the dimensionality, or intervality, of a network before any attempt is made to order the nodes is an important preprocessing step. The higher the intervality, the more confidence can be placed in the ordering obtained by the Fiedler vector. If the intervality is sufficiently small, further eigenvectors of the Laplacian matrix can be used to derive a multidimensional ordering of nodes.

We used the described methods in a socio-economic context, deriving leveling ties of occupations from US census data. The network has a surprisingly high intervality and occupations can be embedded in two

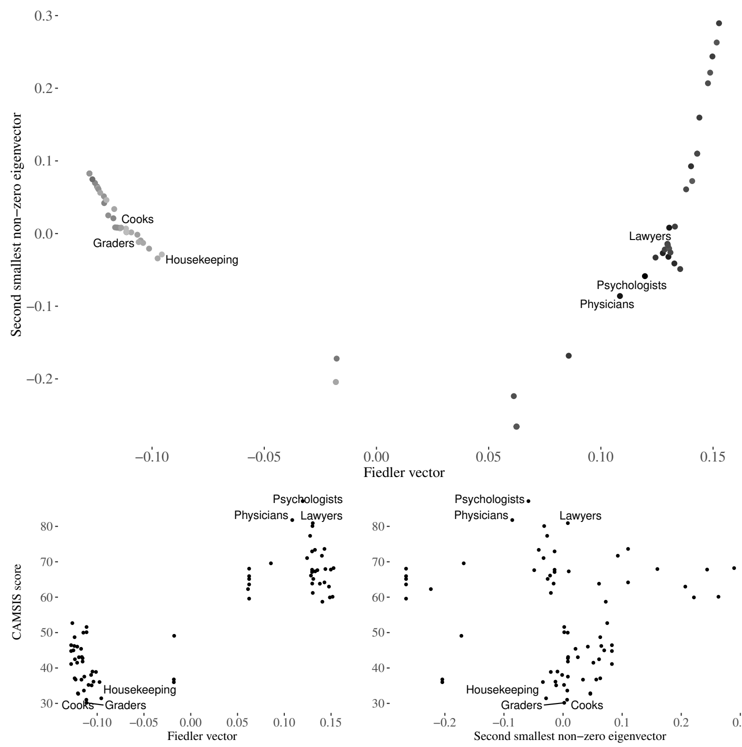


Fig. 6. (Top) Fiedler vector and second smallest non-zero Laplacian eigenvector of the US census network shown in Figure 5. (Bottom) Both vectors separately plotted against the CAMSIS scores.

dimensions. The ordering in at least one dimension roughly coincides with the more established CAMSIS scores. A more thorough empirical examination, however, is required to interpret both dimensions coherently.

So far, we have heavily drawn from existing tools in the literature to analyse leveling ties. Future work should thus focus on the development of a set of new tools, tailored to these networks. To do so, we need to understand the conceptual peculiarities of leveling ties and their prevalence in the literature which ultimately guide the development of appropriate analytic tools.

References

1. Atkins, J., Boman, E., Hendrickson, B.: A Spectral Algorithm for Seriation and the Consecutive Ones Problem. SIAM Journal on Computing **28**(1), 297–310 (1998).
2. Borgatti, S.P., Mehra, A., Brass, D.J., Labianca, G.: Network Analysis in the Social Sciences. Science **323**(5916), 892–895 (2009).

3. Cartwright, D., Harary, F.: Structural balance: A generalization of Heider's theory. Psychological Review **63**(5), 277 (1956)
4. Cattin, M.F., Bersier, L.F., Banašek-Richter, C., Baltensperger, R., Gabriel, J.P.: Phylogenetic constraints and adaptation explain food-web structure. Nature **427**(6977), 835–839 (2004).
5. Cohen, J.E.: Food webs and niche space. Monographs in Population Biology (11), 1—189 (1978)
6. Connelly, R., Gayle, V., Lambert, P.S.: A Review of occupation-based social classifications for social survey research. Methodological Innovations **9**, 2059799116638,003 (2016).
7. Domínguez-García, V., Johnson, S., Muñoz, M.A.: Intervality and coherence in complex networks. Chaos: An Interdisciplinary Journal of Nonlinear Science **26**(6), 065,308 (2016).
8. Erikson, R., Goldthorpe, J.H., Portocarero, L.: Intergenerational Class Mobility in Three Western European Societies: England, France and Sweden. The British Journal of Sociology **30**(4), 415–441 (1979).
9. Fishburn, P.C.: Interval graphs and interval orders. Discrete Mathematics **55**(2), 135–149 (1985).
10. Freeman, L.C.: A Set of Measures of Centrality Based on Betweenness. Sociometry **40**(1), 35–41 (1977).
11. Freeman, L.C.: Spheres, cubes and boxes: Graph dimensionality and network structure. Social Networks **5**(2), 139–156 (1983).
12. Freeman, L.C.: Boxicity and the social context of Swedish literary criticism, 1881–1883. Journal of Social and Biological Structures **9**(2), 141–149 (1986)
13. Fulkerson, D., Gross, O.: Incidence matrices and interval graphs. Pacific Journal of Mathematics **15**(3), 835–855 (1965)
14. Ganzeboom, H.B.G., De Graaf, P.M., Treiman, D.J.: A standard international socio-economic index of occupational status. Social Science Research **21**(1), 1–56 (1992).
15. Garey, M.R., Johnson, D.S.: Computers and Intractability. WH Freeman, New York (2002)
16. Goldberg, P.W., Golumbic, M.C., Kaplan, H., Shamir, R.: Four Strikes Against Physical Mapping of DNA. Journal of Computational Biology **2**(1), 139–152 (1995).
17. Goldthorpe, J.H., Llewellyn, C., Payne, C.: Social Mobility and Class Structure in Modern Britain. Oxford University Press (1980)
18. Golumbic, M.C., Shamir, R.: Complexity and Algorithms for Reasoning About Time: A Graph-theoretic Approach. J. ACM **40**(5), 1108–1133 (1993).
19. Griffiths, D., Lambert, P.S.: Dimensions and Boundaries: Comparative Analysis of Occupational Structures Using Social Network and Social Interaction Distance Analysis. Sociological Research Online **17**(2), 5 (2011)
20. Habib, M., McConnell, R., Paul, C., Viennot, L.: Lex-BFS and partition refinement, with applications to transitive orientation, interval graph recognition and consecutive ones testing. Theoretical Computer Science **234**(1), 59–84 (2000).
21. Heider, F.: Attitudes and Cognitive Organization. The Journal of Psychology **21**(1), 107–112 (1946).

22. Kratochvil, J.: A Special Planar Satisfiability Problem and a Consequence of Its NP-Completeness Discrete Applied Mathematics **52**(3), 233–252 (1994)
23. Lambert, P.S., Bihagen, E.: Stratification Research and Occupation-Based Social Classifications. In: Social Stratification: Trends and Processes, pp. 113–128. Farnham: Ashgate (2012)
24. Laumann, E.O.: Subjective Social Distance and Urban Occupational Stratification. American Journal of Sociology **71**(1), 26–36 (1965).
25. Laumann, E.O., Guttman, L.: The Relative Associational Contiguity of Occupations in an Urban Setting. American Sociological Review **31**(2), 169–178 (1966).
26. Mitchell, J.C., Critchley, F.: Configurational similarity in three class contexts in British society. Sociology **19**(1), 72–92 (1985)
27. Prandy, K., Lambert, P.S.: Marriage, Social Distance and the Social Space: An Alternative Derivation and Validation of the Cambridge Scale. Sociology **37**(3), 397–411 (2003).
28. Quest, M., Wegner, G.: Characterization of the graphs with boxicity $\leqslant 2$. Discrete Mathematics **81**(2), 187–192 (1990).
29. Roberts, F.S.: On the boxicity and cubicity of a graph. In: Recent Progress in Combinatorics, pp. 301–310. Academic Press (1969)
30. Roberts, F.S.: Graph Theory and Its Applications to Problems of Society. SIAM, Philadelphia (1978).
31. Rosengren, K.: The Climate of Literature: Sweden's Literary Frume of Reference, 1953-1976. Lund: Studentlitteratur (1983)
32. Ruggles, S., Genadek, K., Goeken, R., Grover, J., Sobek, M.: Integrated Public Use Microdata Series: Version 7.0 [dataset]. Minneapolis: University of Minnesota, 2017 (2017).
33. Sabidussi, G.: The centrality index of a graph. Psychometrika **31**(4), 581–603 (1966).
34. Schoch, D., Brandes, U.: Re-conceptualizing centrality in social networks. European Journal of Applied Mathematics **27**(6), 971–985 (2016).
35. Stewart, A., Prandy, K., Blackburn, R.M.: Measuring the Class Structure. Nature **245**(5426), 415–417 (1973).
36. Stewart, A., Prandy, K., Blackburn, R.M.: Social Stratification and Occupations. Springer (1980)
37. Stouffer, D.B., Camacho, J., Amaral, L.A.N.: A robust measure of food web intervality. Proceedings of the National Academy of Sciences **103**(50), 19,015–19,020 (2006).
38. Treiman, D.J.: Occupational Prestige in Comparative Perspective. Academic Press, New York (1977)
39. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge University Press (1994)