Identifying Fragments in Networks for Structural Balance and Tracking the Levels of Balance Over Time

Patrick Doreian
University of Ljubljana, Ljubljana, Slovenia &
University of Pittsburgh, PA USA

Andrej Mrvar
University of Ljubljana, Ljubljana, Slovenia

Abstract
This paper presents three items. The first is a brief outline of structural balance oriented towards tracking the amount of balance (or imbalance) over time in signed networks. Often, the distribution of specific substructures within broader networks has great interest value. The second item is a brief outline of a procedure in Pajek for identifying fragments in networks. Identifying fragments (or patterns or motifs) in networks has general utility for social network analysis. The third item is the application of the notion of fragments to counting signed triples and signed 3-cycles in signed networks. Commands in Pajek are provided together with the use of Pajek project files for identifying fragments in general and signed fragments in particular. Our hope is that this will make an already available technique more widely recognized and used. Determining fragments need not be confined to signed networks although this was the primary application considered here.

Keywords: *Signed networks, structural balance, network fragments, temporal balance, and international relations*

Authors
Patrick Doreian, *Research Professor, Faculty of Social Sciences, University of Ljubljana, Ljubljana, Slovenia and Professor Emeritus, Department of Sociology, University of Pittsburgh, Pittsburgh, PA USA.*

Andrej Mrvar, *Professor, Faculty of Social Sciences, University of Ljubljana, Ljubljana, Slovenia.*

Notes
Patrick Doreian and Andrej Mrvar have enjoyed a long collaboration focused on signed social networks. Patrick Doreian formerly edited the Journal of Mathematical Sociology and currently co-edits Social Networks with Martin Everett. Andrej Mrvar is a former editor of Metodološki Zvezki and co-author with Vladimir Batagelj of the award winning program suite Pajek.

Correspondence concerning this work should be addressed to Patrick Doreian, Department of Sociology, University of Pittsburgh, 230 S Bouquet St, Pittsburgh, PA 15213. Email: pitpat@pitt.edu

DOI: http://dx.doi.org/10.17266/35.2.1

1. Introduction

As noted by Taylor (1970), Heider (1946) provided the initial statement of structural balance theory. There have been alternative formulations of 'consistency theories' of signed social relations, including Newcomb (1961), Nordlie (1958), Festinger (1957), Osgood and Tannenbaum (1955) and others (see Abelson *et al*., 1968). However, we use Heider's approach because Cartwright and Harary (1956) provided a formal generalization of his theory, one laying the foundations for analyzing signed social networks in balance theoretic terms. Given temporal data for signed relations, a natural question is how signed network structures change over time regarding balance and how this can be tracked. We demonstrate doing this by using Pajek (Batagelj and Mrvar, 1998). Section 2 provides a brief introduction to the relevant parts of structural balance for our purposes here. The ideas of defining and detecting network fragments are presented in Section 3. The application of fragments by creating specific signed fragments appropriate for measuring balance in signed networks follows in Section 4 where two empirical examples are considered. One is a directed network while the second is undirected. Section 5 extends these ideas so that balance can be tracked in signed networks over time. A summary and suggestions for future work are in Section 6.

2. Structural balance

We consider the initial formulation of balance theory before outlining briefly the blockmodeling and triple counting approaches for measuring imbalance in a signed network.

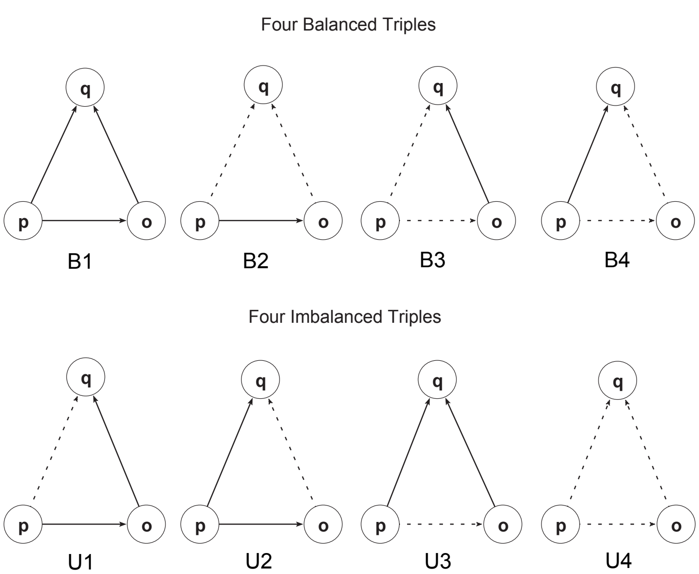

Figure 1: The Eight signed triples in Heider's formulation of structural balance theory.

2.1 The initial formulation

Heider's (1946) approach rests on considering the eight types of triples shown in Figure 1. Positive ties are marked with solid lines while negative ties are marked with dashed lines. One typical signed relation has, as positive ties, 'likes' while negative ties are 'dislikes' for personal relations. The vertices are labeled by p, q and o. The ties are directed as shown by the arrows of the lines. In the top left triple, the ties $p \rightarrow o$, $p \rightarrow q$ and $o \rightarrow q$ are all positive. This was seen as 'balanced' in the sense of there being no discomfort for the three actors. In the second triple, $p \rightarrow o$ is positive with both $p \rightarrow q$ and $o \rightarrow$ q being negative. Both p and o agree by each having a negative tie to q with a positive tie, $p \rightarrow o$. The remaining triples in the top row can be read in the same fashion. When the signs on the three arcs in a triple are multiplied the resulting sign is taken as a measure of the balance of a triple. Triples with a sign of 1 are balanced while triples whose sign is -1 are imbalanced. These triples are shown as Pajek networks in Table 1.

All of the triples in the top row are balanced. We have labeled them B1, B2, B3 and B4 and use these labels in Table 1. They have been expressed in folk aphorisms: The top left triple is captured by "a friend of a friend is a friend" with all ties being positive; the second triple can be viewed as "an enemy of a friend is an enemy" with p seeing o as a friend with both o and p seeing q as an enemy; the third triple can be viewed as "a friend of an enemy is an enemy" with p seeing o as a friend of q when p sees both o and q as enemies; the top right triple is "an enemy of an enemy is a friend" with p viewing o an enemy, seeing o views q as an enemy and p sees q as a friend. All of the triples in the bottom row have a negative sign and are imbalanced. These are labeled U1, U2, U3 and U4 with the labels used also in Table 1. According to Heider (1946), the bottom left triple in the bottom row would be problematic with p seeing q as an enemy while seeing o as a friend but recognizes that o views q as a friend. The other triples can be viewed in a similar fashion. Heider emphasized that balance in a triple induced comfort while imbalance created stress for the actors in such triples.

In a complete network, if all triples are positive, the network is balanced. Empirically, most empirical signed networks are not exactly balanced. This is the case for the empirical networks considered here. The natural methodological issue arising is how to measure the extent to which a signed network is balanced or not balanced. One proposed measure of balance is the proportion of the balanced triples it contains. Its value for a balanced network is 1, the maximum possible value. When some imbalanced triples are present this measure departs from 1. If all triples in a network are imbalanced, the measure takes its lowest value, 0. The question arises: how do we measure the imbalance of signed networks in general? There are two broad approaches: using signed blockmodeling and counting triples.

2.2 Using the line index of balance from a blockmodel

Cartwright and Harary (1956) proved that if a signed network is balanced then the vertices can be partitioned into two subsets such that all of the positive ties are within subsets and all of the negative ties being between subsets. Davis (1967) extended this to *any* number of clusters with the same property of positive ties being located within clusters and negative ties between them. The crucial step for this extension was to define the all-negative triple as balanced. Doreian and Mrvar (1996) observed that the 'structure theorems' of Cartwright and Harary and of Davis implied a blockmodel structure. A positive block is one having no negative ties. In contrast, a negative block has no positive ties. The implied blockmodel of an exactly balanced signed network has positive blocks on the main diagonal and negative blocks elsewhere. As noted above, empirical networks are seldom balanced exactly. When a signed blockmodel is fitted to signed data it provides also a measure of imbalance in the form of the number of ties inconsistent with the relevant structure theorem. In essence, this is the line index of imbalance proposed by Harary, Cartwright and Norman (1965). While the intuitive foundations for blockmodeling are straightforward (Doreian, Batagelj and Ferligoj, 2005), fitting them can be time consuming, especially as the network size increases. Doreian and Mrvar (1996) provided a rapid method for fitting signed blockmodels in Pajek. This line index is one measure of imbalance.

2.3 Using counts of triples to measure imbalance

Another approach is very simple: count the number of signed triples in a signed network. When the triples shown in Figure 1 are counted there are two possible measures of imbalance. The traditional one is the *proportion* of imbalanced triples with the other being the *number* of them (Doreian and Mrvar, 2015). Counting triples is obviously useful when the signed network is complete. However, when signed networks are not complete, this necessitates counting all closed walks and semi-walks (which include triples). Doing this is a non-trivial computational problem. When done, the proportion of imbalanced semi-walks, most likely, would depart from the corresponding measure using only triples. The obvious question is whether this matters. We think it does not. The core substantive ideas of Heider are formulated in terms of triples. This suggests counting triples is more appropriate for socio-psychological processes than counting the longer semi-walks. How can this be done simply? An effective way of counting triples can be achieved by using the concept of fragments. The general approach is to define fragments of specific forms, identify them and count them. Doing this is achieved straightforwardly in Pajek (see Batagelj and Mrvar (1998)). Our focus here is on signed triples as fragments, an idea described in Section 3.

We note that, hitherto, the line index and proportion of imbalanced cycles as measures of imbalance have been closely related: they ought to tell the same story. As the size of the networks we can study has increased dramatically in recent decades, the blockmodeling approach is likely to be less useful due to the computational complexities involved. For these larger networks, counting triples will be a practical alternative.

3. Fragments

Characterizing networks when they are large has posed problems. One strategy is to consider carefully constituent parts of networks. As a result, researchers have been interested in identifying such smaller parts of larger networks having special properties (characteristic shapes) across multiple fields. Such smaller parts are called fragments, patterns, or motifs. Fragment searching was first implemented in Pajek in 1997. (See also Milo, Shen-Orr, Itzkovitz S., Kashtan N, Chklovskii D., and Alon, U (2002) for a discussion of motifs from the perspective of physicists approaching network analysis.) Fragment (pattern) searching is a general approach for investigating the structure of large complex systems. Frequencies and locations of such interesting fragments often provide short descriptions of network structures in terms of the distributions of well-defined fragments contained in them. This could include cycles, k-stars and cliques of any size. Given an interest in structural balance, defining the eight triples in Figure 1 as fragments is a natural step. Doing this sets up the use of fragment searching for all signed triples.

We provide a simple example of fragment searching in Section 3.1 following some remarks on this topic. See also de Nooy, Mrvar and Batagelj (2011). A general backtracking algorithm is applied for fragment searching. Several applications have shown that if the selected fragments do not occur too frequently in a large sparse network, the algorithm is extremely fast. It can be applied to very large sparse networks.

Fragment searching in networks was first applied to large molecules in chemistry (e.g. DNA), for searching for carbon rings and other structures. Later, fragment searching was successfully applied to searching for relinking families through marriages in genealogies. Every semi-cycle found in a p-graph representation of a genealogy represents some relinking (via blood or non-blood ties) through marriages. See White, Batagelj and Mrvar (1999), Batagelj and Mrvar (2008).

3.1 A simple motivating example

Consider the small unsigned network in Figure 2. Suppose the task is to identify all 3-cycles and 4-cycles as fragments. Some clear 3-cycles - $\{a, b, c\}$ and $\{o, p, q\}$ - are marked with green edges. Some clear 4-cycles – $\{e, f, g, h\}$, $\{k, l, m, n\}$ and $\{q, r, s, t\}$ – are marked with blue edges. The subgraph involving the vertices i, j, k and t is a little more complicated. There is a 4-cycle involving all four of them. The relevant edges are marked in maroon. Note there are also two 3-cycles in this subgraph – $\{i, j, k\}$ and $\{j, k, t\}$. The (j, k) edge is unambiguously part of the two 3-cycles and is marked in green. The edges marked in maroon are each in a 4-cycle and a 3-cycle. By a visual inspection, there are four 3-cycles and four 4-cycles. However, such visual examinations have no practical value when searching larger networks having hundreds or thousands of vertices. A systematic and practical procedure is required.

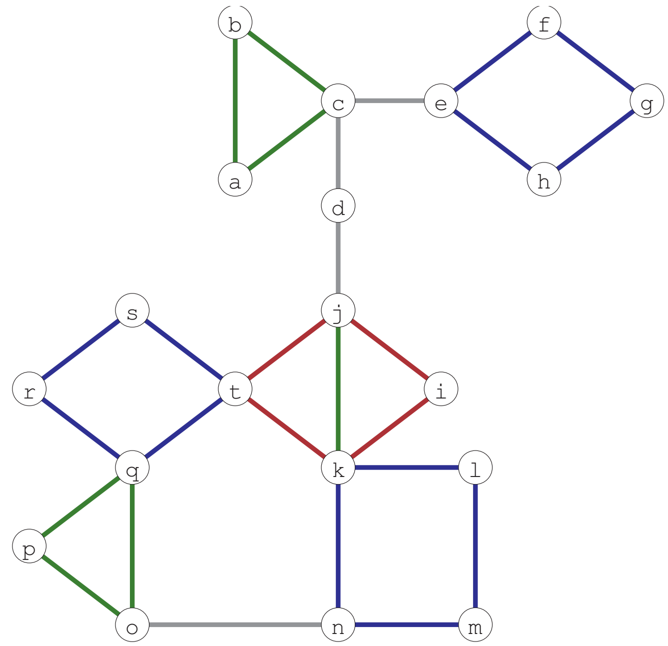

Figure 2: An undirected graph with 3-cycles and 4-cycles to be identified and counted.

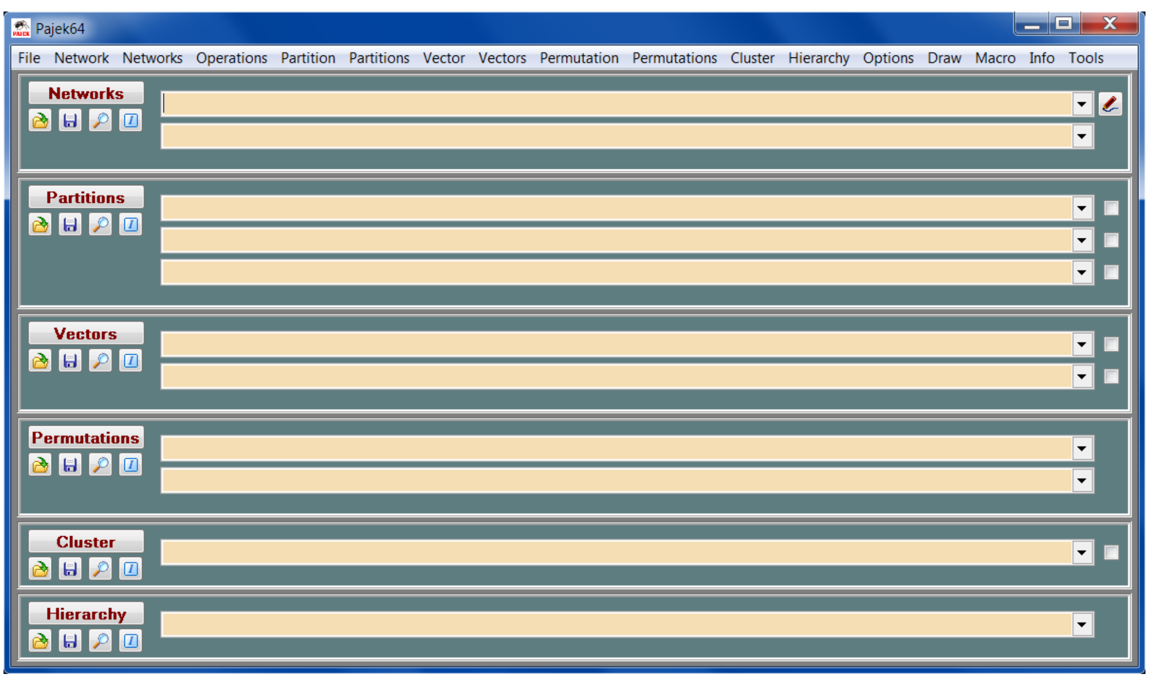

Figure 3. The main window for using Pajek

Pajek provides such a practical method. Figure 3 shows the main window for Pajek when it is run. Across the top of this window is main menu containing items: File, Network, Networks, Operations, Partition, Partitions, Vector, Vectors, Permutation, Permutations, Cluster, Hierarchy, Options, Draw, Macro, Info and Tools. Checking on any of these opens a dialogue box. Each dialogue box has its own set of relevant objects and operations. The primary one we use in the following is Networks (which as the name implies leads to working with multiple networks) because fragments are defined as networks. We search for such fragments in a larger network. This requires two networks. In the simple example of Figure 2, the network within which the search is done is the one in the figure and a 3-cycle would be a fragment for which a search is done. When there are searches for multiple fragments, each fragment has its own search. (We note that the Draw option was used to draw all of the network diagrams we show.) Much fuller descriptions can be found in the Pajek manual and in de Nooy, Mrvar and Batagelj (2011).

There are six horizontal panels in the main body of the window for Networks, Partitions, Vectors, Permutations, Cluster and Hierarchy. Under each of these names are some icons. Reading from the left the icons are used to read Pajek files, save Pajek files, view or edit a file that has been read (or created in Pajek) and obtain information about that file. We use the Networks horizontal panel for identifying and counting fragment types. There are horizontal lines in this panel and two are used for fragment searching as described below. The first line will contain the fragment with the second containing the network within which the search is done.

4. Two Empirical Signed Networks

4.1 Analyses for a directed signed network

Section 4.1.1 focuses on the global-level analysis of counting all of the signed triples in an empirical network and presenting the results. In addition, given such distributions, it is natural to ask about the involvement of specific or all egos in them. This is considered briefly in Section 4.1.2 before returning to the measurement of imbalance in signed networks.

4.1.1 Counting all triple types

We demonstrate doing this for larger and/or more complex networks by identifying and counting the triples of Figure 1 as the fragments to be identified in the network shown in Figure 4. The data come from Lemann, and Solomon (1952). Women in a college were asked about signed preferences about whom the women would like or not like to do activities. Each woman was asked to name others but did not know the preferences of anyone else in their group. The data for all four relations were reanalyzed from a blockmodeling perspective (Doreian, 2008).

The data used here feature one relation (going on a double date) as shown in Figure 4. Blue lines represent positive ties and red lines show negative ties. Some positive and some negative ties are reciprocated. For visual simplicity, pairs of positive reciprocated arcs are represented by solid blue edges rather than by two arcs. Pairs of negative reciprocated arcs are represented by dashed red edges rather than by two arcs. Remarkably, some reciprocated pairs have opposite signs (e.g. *m-r* and *j-s*). There are 21 vertices with 63 positive arcs and 63 negative arcs.[1] We provide these data in the zipped file available at: http://mrvar.fdv.uni-lj.si/pajek/SVG/CoW/cow.zip.

Table 1 shows each of the eight triples in Figure 1 written out as Pajek network files. These define a set of fragments that were contained in a Pajek project file (see below) to facilitate fragment searching. This project file is provided in the zipped file as well.

The first listed network (as a fragment) has three lines at the top started by *: a signal to Pajek as to how the line in the file is to be read. Having the * start these lines is mandatory. The first line gives its name 'Network Only Positive', the second gives the number of vertices (there are three) and the third gives the type of lines in the network (arcs). The next three lines contain the data with positive ties. The first network in the second column is the first imbalanced triple (bottom left in Figure 1). There are three imbalanced triples with a single negative tie. We keep them distinct by giving them different names as they are searched for separately.

[1] Given the data were collected a long time ago it is not surprising the data collection design was fixed choice for all actors. While there are drawbacks with this design, they are irrelevant for demonstrating the signed fragments analysis.

Balanced Triples	Imbalanced Triples
*Network Only Positive (B1)	*Network One Negative /1 (U1)
*Vertices 3	*Vertices 3
*Arcs	*Arcs
1 2 1	1 2 1
1 3 1	1 3 -1
2 3 1	2 3 1
*Network Two Negative /1 (B2)	*Network One Negative /2 (U2)
*Vertices 3	*Vertices 3
*Arcs	*Arcs
1 2 1	1 2 1
1 3 -1	1 3 1
2 3 -1	2 3 -1
*Network Two Negative /2 (B3)	*Network One Negative /3 (U3)
*Vertices 3	*Vertices 3
*Arcs	*Arcs
1 2 -1	1 2 -1
1 3 -1	1 3 1
2 3 1	2 3 1
*Network Two Negative /3 (B4)	*Network Only Negative (U4)
*Vertices 3	*Vertices 3
*Arcs	*Arcs
1 2 -1	1 2 -1
1 3 1	1 3 -1
2 3 -1	2 3 -1
The labels B1, B2, B3 and B4 are the same as in Figure 1. (The secondary labels /1, /2 and /3 are for the three types of triples having two ties that are negative.)	The labels U1, U2, U3 and U4 are the same as in Figure 1. (The labels /1, /2 and /3 are for the three types of triples having one tie that is negative.)

Table 1: The eight network fragments defined by the triples in Figure 1.
Note: The eight fragments are stored after each other in one column in Pajek, not in two panels.

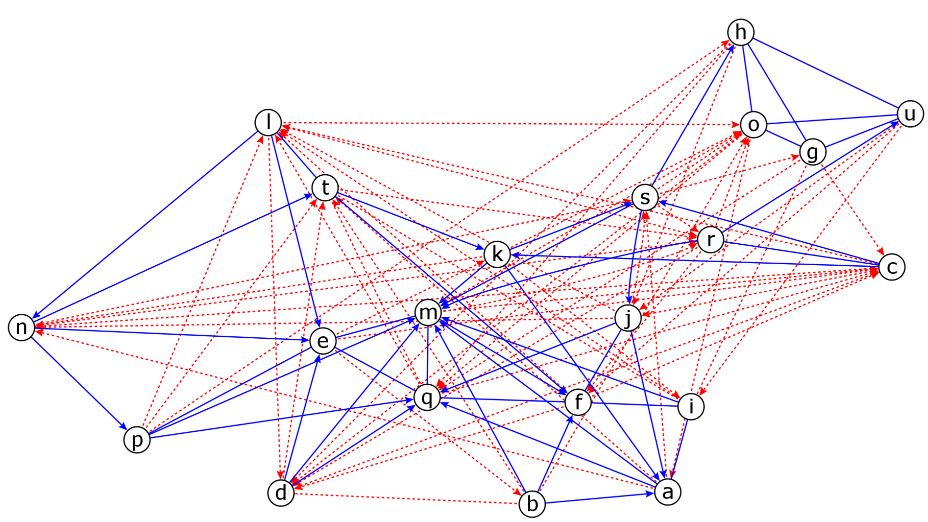

Figure 4. A Directed Signed Relational Network
Notes: Blue solid lines - positive relation; red dashed lines - negative relation

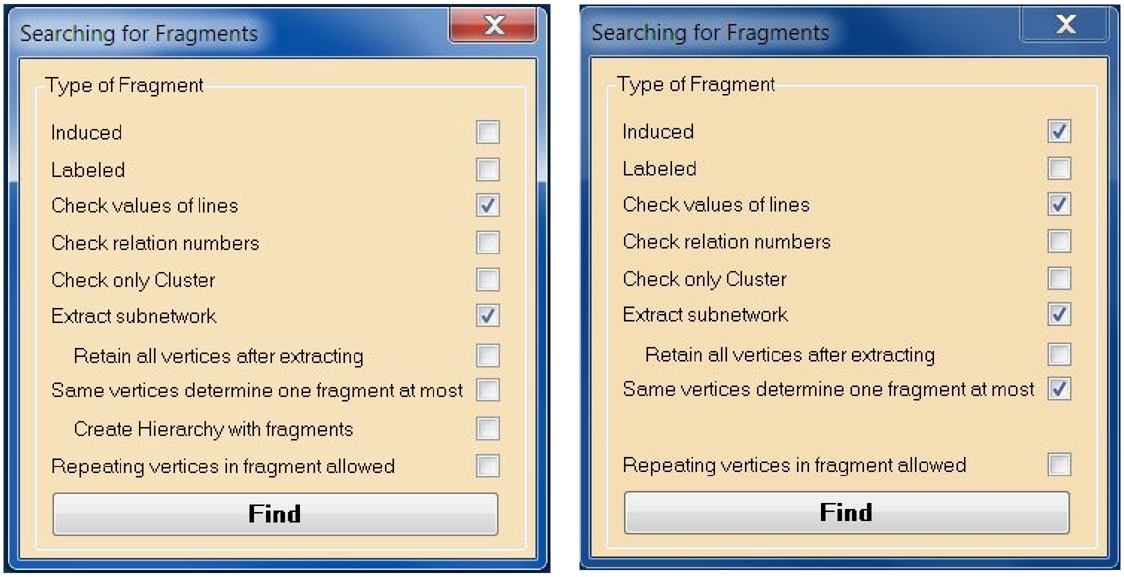

Table 2. Options for the searching for fragments procedure
Notes: The dialogue box on the left shows the options for obtaining the triples of Table 1 as fragments in the network shown in Figure 4.

A particularly useful feature of Pajek is the option for saving Pajek project files. At any stage of an analysis, most usefully at a provisional end, all of the objects that have been defined can be saved to a project file with a single command. This project file can then be read subsequently by Pajek. Clicking on File in the top row of Pajek's main window opens a dialogue box in which two of the options are for saving and reading Pajek project files which will have the extension paj. For fragment identification, the networks defined by the fragments can be stored in such a project file. This means the fragments can be defined once and then recalled for each new analysis. For structural balance with arcs, the eight networks defined by the triples of Figure 1 are stored in a project file as shown in Table 1. The project file for them has a single column rather than the two shown in the table. It is provided in the zipped file. The starting * for each fragment is read by Pajek as signaling a new fragment. When the file is read, all of the fragments are read but each fragment search is done separately.

Balanced		Imbalanced	
Triple type	Count of triples	Triple type	Count of triples
Only positive	65	One Negative/1	8
Two Negative/1	57	One Negative/2	6
Two Negative/2	25	One Negative/3	21
Two Negative/3	22	One Negative	20

Table 3: Counts of signed triples for the directed signed network

The network file in which the searching is done has a similar structure but will be much larger with 21 vertices, and each is listed on a separate line with the network ties also listed on separate lines. The steps in Pajek for extracting these types of triples and counting them are as follows:

Getting the data into Pajek:

- Read the data from a network file (*.net) with * replaced by the file's name.[2]
- Read the Pajek project file (with the form (*.paj). (We labeled this as balancefragarcs.paj as it contains the eight triple types for use with any signed directed network. The label name is arbitrary.)

Mobilizing Pajek to determine and count the types of triples

- Select the each fragment type (one at a time) from the Pajek project file as the first network.
- Select the network data as the second network.
- Check Networks from the top menu bar in Pajek (to open a dialogue box).

[2] There are two conventions for use a * that need to be kept distinct. The one described thus far for Pajek files in this paper is internal to Pajek. The second convention is for all files when it is used a token for any name of a file. Users are free to choose their own names for files. The network with the data of figure 4 was called haddate1.net. Again, the file name is arbitrary. The data are provided in the zipped file. We recommend strongly that Pajek users upgrade always to the most recent version. When this paper was finalized. The Pajek version was 4.04. Older versions required that there be no spaces in Pajek file names. This is no longer the case.

- Check ‘Fragments (First in Second)’ in this dialogue box. This opens another small dialogue box with the options shown in Table 2.
- Select the appropriate options.
- Check Find

Table 2 shows two alternatives for selecting options for finding fragments. In the left panel, the options we selected for completing the above analysis were: ‘Check values of lines’ and ‘Extract subnetwork’. These choices are the checked boxes in the left panel of Table 2. They are appropriate for the directed network in Figure 4. An alternative set of options is shown in the right panel. When these options are used for obtaining the signed fragments in Figure 4 the outcomes are incorrect. However, both sets of options led to the same outcomes for the undirected network in Figure 2. Care is needed in selecting options as they can produce different outcomes depending on the structure of the network and the goals of the analysis.

Directed signed networks present problems for using the right hand set of options when there are reciprocated ties present. This holds regardless of whether these ties have the same sign or different signs. The directed network in Figure 4 has such dyads. When checking ‘Induced’ in Fragment options, such triples are not counted as correct triads (since there are additional arcs not only the ones needed for fragment). For undirected networks, this issue does not arise.

The number of fragments when Find is checked in the fragment dialogue box will appear in the output appearing on the screen. The counts of the signed fragments are shown in Table 3. At face value, the proportion of balanced triples for Table 3 is 169/224 = 0.75 indicating more balance than imbalance. However, the blockmodel analysis of Doreian (2008) showed there were more than two positions (clusters) of the actors. This value for imbalance is not appropriate. Fortunately, the Davis (1967) formulation provides a solution.

As noted above, the Davis generalization of the initial structure theorem of Cartwright and Harary (1956) dealt with networks having more than two positions by defining it as balanced. Indeed, a strong general case can be made for not considering it as imbalanced when there are more than two positions. When this is done, the measure of balance is 189/224 = 0.84, more consistent with the small line index of balance obtained from the signed blockmodel. We note that the blockmodel fitted according to the method presented in Doreian and Mrvar (2009) contains a negative diagonal block.

4.1.2 Ego-level properties for fragments

Given the primary purpose of this paper, the above task completion is enough. But having identified fragments, a natural avenue of inquiry is to think about specific vertices being located within fragments. In addition to the global result of counted fragment types, Pajek also provides ego-level results. Table 3 shows, among other things, there are 21 triples of the type labeled ‘One Negative/3’. We can ask about the frequency with which the vertices are present in this type of triple. The results are shown in Table 4. In terms of frequencies, vertex m heads the list by belonging to eight of these triples. Vertex f comes next with six followed by d, e, and t each with five. At the other extreme, vertices c, j and n belong to none. The sum of fragment counts in the bottom row of Table 4 is 63 (there are altogether 21 triples, each contains 3 vertices).

The same type of analysis can be done for each of the types of triple. As one further example, consider membership in the all-negative triples, labeled ‘Only negative’. The results are shown in Table 5. This time, vertex d heads the list with nine such memberships followed by vertices l and r with seven of them. There are six vertices having no involvement in all-negative triples. Of some interest is vertex m which had the highest count for the triples in Table 4 but is very low regarding the memberships in all-negative triples.

It seems clear that the place of specific vertices can be assessed through the types of triples defined in terms of structural balance. This suggests ways of coupling global features of a signed network expressed in triple types and the involvement of egos in them.

Vertex	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u
Fragment Count	2	4	0	5	5	6	1	2	4	0	3	3	8	0	2	1	4	3	4	5	1

Table 4. The distribution of the number of times vertices are present in ‘One Negative/3’ triples

Vertex	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u
Fragment Count	1	3	6	9	0	4	1	0	0	3	0	7	1	4	6	0	2	7	4	2	0

Table 5. The distribution of the number of times vertices are present in ‘Only Negative’ triples

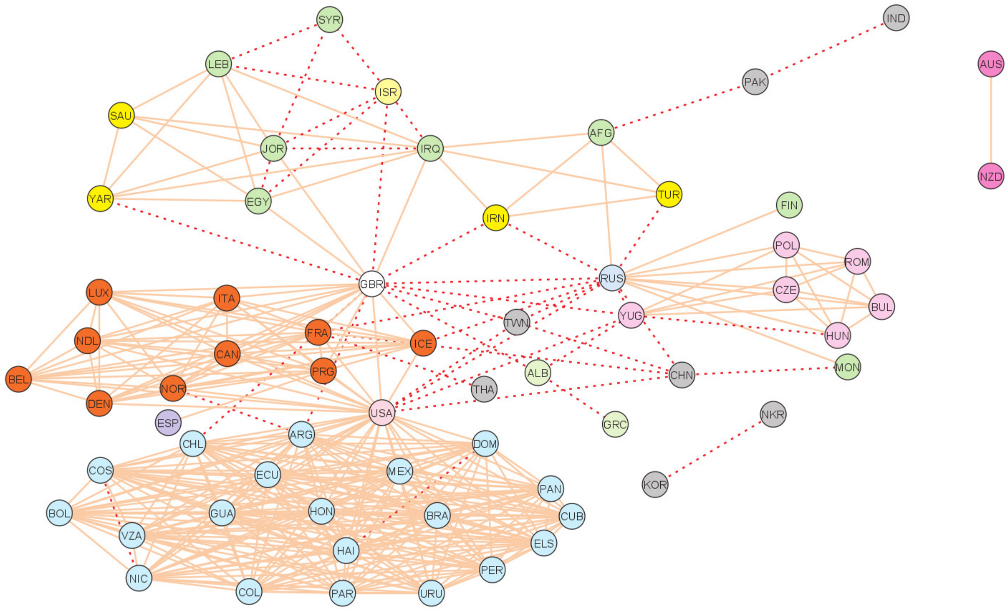

Figure 5. The CoW signed undirected network for 1946-49
Notes: Solid lines represent positive ties with dashed lines representing negative ties

4.2 Measuring imbalance in an undirected network

We return to the main theme of this paper by showing how to measure imbalance in signed networks using a much larger network with undirected ties. This larger network introduces some new methodological issues as discussed in Doreian and Mrvar (2015).

This undirected signed network has 64 vertices and 362 edges. The vertices represent countries which are linked by positive and negative ties. This example is taken from the Correlates of War (CoW) project for nations in the world system following WWII.[3] This network is for the period 1946-1949. The positive ties are for joint memberships in alliances, unions and inter-governmental agreements. The tie is binary. The negative ties are for being at war, in conflict with each other without military involvement, border disputes and sharp ideological or policy disagreements. When there is a negative tie between states that otherwise have a positive tie, the negative tie is used. There are 320 positive edges and 42 negative edges. In contrast to the small network in Figure 4, there is a major difference in the number of positive and negative ties. While this raises some issues regarding the measures of imbalance, the counting of signed triples is not affected. The network is shown in Figure 5. The design of the layout reflects a blockmodel fitted according to balance theoretical ideas. The colors of the vertices represent membership in clusters (positions) determined by blockmodeling (see Doreian and Mrvar, 2015 for details).

When there are only edges in the signed network, the counting of fragments takes the form of counting signed 3-cycles with edges. The number of possible signed 3-cycle types is four. There are only two balanced and two imbalanced 3-cycles. Considering the triples in the top panel of Figure 1, but with edges instead of arcs, there is the all positive 3-cycle. All of the other triples with edges have the same structural form with two negative edges. From the bottom panel, an all negative 3-cycle is present and the other three imbalanced triples with edges have the same structural form with one negative edge. As a result, the Pajek project file with the fragments has the four networks shown in Table 6. The actual project file used is provided in the zipped file. The steps required for obtaining these signed 3-cycles are the same as for the first example only a different Pajek project file with only these four types of 3-cycles was used with the fragments. The relevant counts are shown in Table 7.

The blockmodel shown in Figure 5 through the coloring of the vertices makes it clear there are more than two positions in the world system. The appropriate measure of balance, given the Davis (1967) formulation, is 0.966 (1593/1656). Were the all-negative triple considered as imbalanced, the measure of balance would

[3] See http://www.correlatesofwar.org/. The data were provided kindly by Daniel Halgin of the Links program at the University of Kentucky. They were constructed as part of a larger project on signed networks. We appreciate greatly his generosity.

be 0.954 (1579/1656).[4] Either way, the signed network of nations following WWII was highly balanced. By itself, this is of modest interest. Whether balance (or imbalance) changed over time has greater interest value. Even more important is how – and why - the measures of balance/imbalance changed over time. We consider this in Section 5.

5. Measuring and tracking imbalance through time

The larger data set has signed networks for 51consecutive time points. The network expanded from having only 64 nations to a maximum size of 155 because new nations joined the world system through gaining independence or being formed through dissolution of states, especially the USSR and the former Yugoslavia. Nations having few ties can drop out when these ties are severed or can join the international network when new ties involving them are created.

In terms of identifying fragments and counting them, the above procedure is repeated for each time point. However, as the same commands are going to be repeated for every time point the procedure can be made more efficient by using Pajek's 'Repeat last command' feature. The modified Pajek commands are:

- Load all 51 networks and 4 fragments in Pajek. This can be done most easily by having all of the networks stored in a Pajek project file as they will be loaded in one step.
- Select the first fragment (from the fragments project file) as the First Network.
- Select the first loaded network as the Second Network.
- Run Fragments searching as described above.
- Check 'Repeat Last Command'. (This will open another small dialogue box with further options appearing.)
- Check the 'Fix (First) Network' button in this dialog box. Doing this sets Pajek up to search for the same fragment in all of the rest of networks, starting with the second network.
- Check the 'Repeat Last Command' button and, when asked for the number of repetitions, enter 50 (as there are 50 more networks for which first fragment should be found). In general, this entered number will change depending on the number of networks in which fragments are sought. Pajek then searches for this fragment in all of the networks that were loaded.
- To search for other types of fragments (as loaded from the fragment Pajek project file) in all of the networks (i.e. in this case all 51 time points) execute the above sequence of commands for each of the remaining fragments.

Balanced 3-cycles	Imbalanced 3-cycles
*Network Only Positive Undirected *Vertices 3 *Edges 1 2 1 1 3 1 2 3 1	*Network Only Negative Undirected *Vertices 3 *Edges 1 2 -1 1 3 -1 2 3 -1
*Network Two Negative Undirected *Vertices 3 *Edges 1 2 1 1 3 -1 2 3 -1	*Network one Negative Undirected *Vertices 3 *Edges 1 2 -1 2 3 1 1 3 1

Note: The four fragments are stored after each other in one column in Pajek, not in two panels
Table 6. The four network fragments defined by 3-cycles

[4] The corresponding measures of imbalance are 0.034 and 0.046 respectively.

Balanced		Imbalanced	
Triple type	Count of triples	Triple type	Count of triples
Only positive edges	1556	One negative edge	63
Two negative edges	23	Only negative edges	14

Table 7. Counts of signed 3-cycles for the undirected signed network

One additional result of the 'Repeat Last Command' is a Vector called 'Number of Fragments' in which counts of all fragments for all 51 time periods are stored. The counts of fragments can be stored and organized by the time points. Figure 6 shows the results of computing the proportions of balanced signed triples for each time point. Clearly, there were huge changes in the levels of balance (and imbalance depending how the system is viewed.) Of course, the next task for a broader project is to account for the changes by coupling them to events taking place in the world over time. This will be no easy task but, for our purposes here, the task of tracking changes in balance has been completed. A preliminary effort at doing this is contained in Doreian and Mrvar (2015).

Figure 6 makes it clear that the level of imbalance varied greatly over time. One of the empirical hypotheses espoused by structural balance theorists was that signed networks moved towards a balanced state. Even with the early small group datasets that were examined, the empirical evidence tended not to be consistent with this hypothesis. See, for example, Doreian et al. (1996) and Hummon and Doreian (2002). The hypothesis of movement towards balance, even though Heider's initial formulation made it seem very plausible, was not a fruitful longer-term hypothesis for the field: it led to a one-directional view regarding change in signed networks towards balance and obscured a more important question. See Doreian and Mrvar (2015). A more fruitful line of inquiry is to ask about the conditions under which signed systems do move towards balance and the conditions under which they move away from balance.

Another issue addressed by Doreian and Mrvar (2015) is the utility of the proportion of balanced triples when the number of positive ties far exceeds the number of negative ties. The same issue would occur if the number of negative ties exceeded the number of positive ties by a wide margin. This issue has been obscured by the ways in which data have been collected hitherto. For a comparison with the blockmodeling approach when there are disproportional numbers of signed ties, we argue the best comparison is between the *number*[5] of imbalanced triples and the line index of imbalance. The temporal plot of imbalance using this measure is shown in Figure 7. The correlation between the line index of imbalance and the number of imbalanced triples is 0.91. The contrast between the trajectories shown in Figure 6 and 7 is discussed further in Doreian and Mrvar (2015).

While there is no reason to expect a perfect correspondence between these two measures, they are tracking something in a similar fashion. One reason for the slight difference between these measures is that the line index is constructed for the networks as a whole while counting triples or 3-cycles is far more local. Longer cycles and semi-walks are not counted when only triples are considered. However, as noted above, for studying the balance theoretic dynamics of signed networks, such longer fragments have far less utility. For a blockmodeling analysis, the line index is preferable as it is integral to the delineation of the blockmodel structure of a signed network. But if all that is required is a useful measure of the change in balance of a signed network, then using triples provides a simple measure that is far easier to use than conducting blockmodeling for as many networks as were considered for the CoW data.

Regardless of how this issue of the difference between using proportion or number of imbalanced triples to measure imbalance is resolved, the role of counting fragments is clearly useful. The debate will be one of examining the relative merits of the number of imbalanced triples and the proportion of them and will depend on the signed networks that are studied. This will hinge on the relative number of positive and negative ties in the signed network.[6]

6. Conclusion and extensions

A simple method for tracking change in the levels of structural balance (or imbalance) of a signed network was

[5] A very strong case can be made for discounting the presence of the all-positive triples as they are the most frequent type of triple in these data. The real interest in terms of balance theoretic phenomena involves the negative ties and how they are changed over time.
[6] For readers interested in analyzing the temporal data, they can be found at the following link:
http://mrvar.fdv.uni-lj.si/pajek/SVG/CoW/. A zipped version containing all of the data files can be found at: http://mrvar.fdv.uni-lj.si/pajek/SVG/CoW/cow.zip. The data file for the network shown in Figure 5 is N46.net. The other networks for successive periods are numbered in a similar fashion.

presented for signed networks through using fragment identification and then counting them. If needed, this approach can be applied to much larger networks than the examples presented here. Of equal importance, the idea of fragments is far more general than this application for signed networks might suggest. Our hope is that it will be used more often in social network analysis as a way of characterizing network structure depending on the substantive concerns of researchers.

There is another possible way in which imbalance could be measured. It is the number of vertices whose removal creates a balanced network. This is also an NP-hard computational problem. It may be that the distributions of vertices in imbalanced triples could be used to address this issue. Also, further exploration of coupling of global imbalance with the local imbalance at vertices merits further attention.

The data for the network in Figure 5 is the first of a sequence of 51 networks. Over time, this network expanded considerably. This raises the issue of considering the impact of changing system size on the number of fragments present in the network. Doreian and Mrvar (2015) showed that the number of signed triples with negative edges did not expand greatly save for the all-positive triples which exploded. This was due to the expansion of positive ties which were far less costly

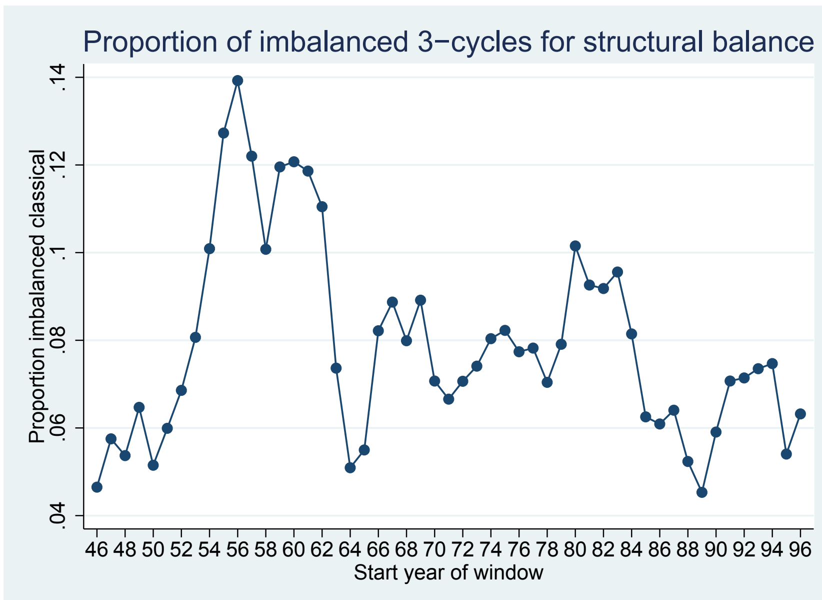

Figure 6. Tracking the proportion of imbalanced triples through time for a changing network.

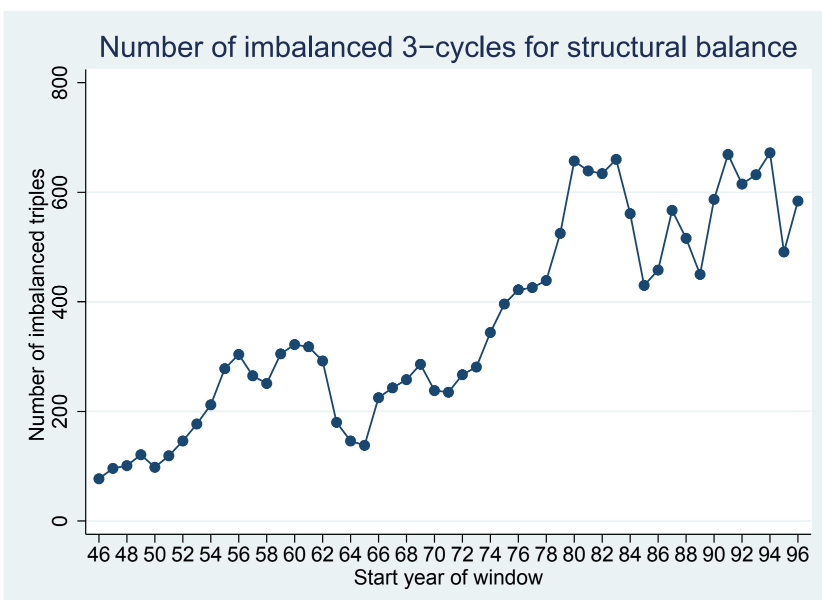

Figure 7. The number of imbalanced 3-cycles over time for the CoW data.

to maintain than the negative triples for countries. The impact of extreme disproportions in the numbers of positive and negative ties may be more consequential than changes in the number of units in the network.

Regarding balance, a more general problem occurs when signed networks (or their closest to balance form) do not conform to the blockmodel structure implied by the structure theorems. The CoW data have positive blocks off the main diagonal and negative blocks on the main diagonal of the blockmodel. The coloring of the vertices in Figure 5 comes from a blockmodel fitted according to relaxed structural balance (Doreian and Mrvar, 2009). Within the blockmodeling approach to signed networks there are two general concerns. One is the delineation of the blockmodel structure with the other being establishing a measure of imbalance for the network as a whole. The correspondence between the line indices from such blockmodels and the counts of triples has not been explored. Examining this relationship is ongoing. However, for signed networks with forms close to those anticipated by the structure theorems, using measures based on identified triples and 3-cycles will be fully appropriate. They may be more practical measures of balance or imbalance if blockmodeling becomes impractical or too time consuming. Moreover, their simplicity makes them easy to interpret.

References

Abelson, R.P., McGuire, W. J., Newcomb, T. E., Rosenberg, M. J. and Tannenbaum, P.H. (Eds.) (1968): Theories of Cognitive Consistency: A Sourcebook, Chicago: Rand- McNally.

Batagelj, V. and Mrvar, A. (1998): "Pajek - A Program for Large Network Analysis", Connections, 21(2), 47-57.

Batagelj, V. and Mrvar, A. (2008): "Analysis of kinship relations with Pajek", Social Science Computer Review, 26(2), 224-246.

Cartwright, D. and Harary, F. (1956): "Structural balance: A generalization of Heider's theory", Psychological Review, 63: 277-292.

Davis, J. A. (1967): "Clustering and structural balance in graphs", Human Relations, 20: 181-187.

de Nooy, W., Mrvar, A., and Batagelj, V. (2011): Exploratory Social Network Analysis with Pajek: Revised and Expanded. 2nd Edition. New York: Cambridge University Press.

Doreian, P., (2008): "A Multiple Indicator Approach to Structural Balance", Social Networks, 30: 247-258.

Doreian, P., Batagelj, V. and Ferligoj, A. (2005): Generalized Blockmodeling. New York: Cambridge University Press.

Doreian, P., Kapuscinski, R., Krackhardt, D. and Szczypula, J. (1996): "A brief history of balance through time", Journal of Mathematical Sociology. 21: 113-131

Doreian, P. and Mrvar, A. (1996): "A partitioning approach to structural balance", Social Networks. 18: 149-168.

Doreian, P. and Mrvar, A. (2009): "Partitioning signed social networks", Social Networks, 31: 1-11.

Doreian, P. and Mrvar, A. (2015): "Structural balance and signed international relations", Journal of Social Structure, Volume 16 Article 2.

Feld S. L. and Elsmore, R. (1984): "Patterns of Sociometric Choices: Transitivity Reconsidered." Social Psychology Quarterly, 45: 77-85.

Festinger, L (1957). A Theory of Cognitive Dissonance, Evanston, IL: Row, Peterson and Co.

Harary, F., Norman, R. Z. and Cartwright, D. (1965): Structural Models. New York: John Wiley and Sons.

Heider, F. (1946): "Attitudes and cognitive organization", Journal of Psychology. 21: 107-112. 58: 5-21.

Hummon, N. P. and Doreian, P. (2003) "Some dynamics of social balance processes: Bringing Heider back into balance theory", Social Networks. 25: 17-49.

Lemann, T.B. and Solomon R. L. 1952. "Group Characteristics as Revealed in Sociometric Patterns and Personality Ratings". Boston: Beacon House, Sociometry Monographs, 27.

Milo R, Shen-Orr SS, Itzkovitz S, Kashtan N, Chklovskii D., and Alon U (2002): "Network motifs: simple building blocks of complex networks". Science 2002, 298:824-827.

Newcomb, T. M, (1961). The Acquaintance Process. New York: Holt, Rinehart and Winston

Nordlie, P. (1958): A Longitudinal Study of Interpersonal Attraction in a Natural Setting, Unpublished Ph. D. Dissertation, University of Michigan.

Osgood, C.E. and Tannenbaum, P.H. (1955). "The principle of congruity in the prediction of attitude change". Psychological Review, 62: 42-55

Taylor, H. F. (1970): Balance in Small Groups. New York: Van Nostrand Reinhold. Theil, H. (1967): Economics and Information Theory. Chicago: Rand McNally.

White, D. R., Batagelj, V., and Mrvar, A. (1999): "Analyzing Large Kinship and Marriage Networks with Pgraph and Pajek". Social Science Computer Review, 17(3), 245-274.