4 Sentiments and friendship

4.1 Introduction

In the preceding chapter, we discussed several techniques for finding cohesive subgroups within a social network. People who belong together tend to interact more frequently than people who do not. In the current chapter, we extend this idea to affective relations which are either positive or negative, for instance, friendship versus hostility, liking versus disliking. We expect positive relations to occur within subgroups and negative relations between subgroups.

Hypotheses about patterns of affective relations stem from social psychology and they are widely known as balance theory. First, we will introduce this theory and discuss how it was incorporated in network analysis. Then, we will apply it to affective relations, that is, social relations which are subjective and mental rather than tangible.

4.2 Balance theory

Social psychology is interested in group processes and their impact on individual behavior and perceptions. In the 1950s, Fritz Heider formulated a principle which has become the core of balance theory, namely, that a person feels uncomfortable when s/he disagrees with his or her friend on a topic. Figure 1 illustrates this situation: P is a person, O is another person (the Other), and X represents a topic or object. P likes O, which is indicated by a positive line between P and O, but they disagree on topic X because P is in favor of it (positive line) whereas O is opposed to it (negative line). Note the convention of drawing negative relations as dashed lines.

Heider predicted that P would become stressed and feel an urge to change the imbalance of the situation, either by adjusting his opinion on X, by changing his affections for O, or by convincing himself or herself that O is not really opposed to X. Research in small groups corroborated the hypothesis that people feel stressed in a situation of imbalance.

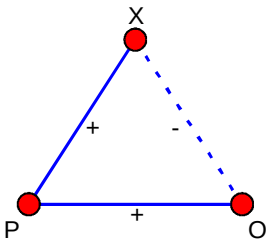


Figure 1 - A Person - Other - Object (X) triple.

A social psychologist (Dorwin Cartwright) and a mathematician (Frank Harary) translated Heider's ideas into network analysis. They defined a special kind of

network to represent structures of affective relations, namely a signed graph. In a signed graph, a positive or negative sign is attached to each line indicating whether the associated relation, e.g., an affection, is positive or negative.

A **signed graph** is a graph in which each line carries either a positive or a negative sign.

In a signed graph, Heider's Person - Other - Object triple is represented by a cycle, that is, a path in which the first and last vertex coincide. All balanced cycles contain an even number of negative lines or no negative lines at all, for instance, there is one negative line in the cycle of Figure 1, which is an uneven number, so this triple is not balanced. P, and possibly O, will feel stressed in this situation.

However, affective relations do not need to be symmetrical. My feelings for you may differ from your feelings towards me. Affections are 'projected' from a person to something or someone else. Therefore, it is usually better to represent affect relations by arcs rather than edges. It is easy to generalize balance theory to signed directed graphs: ignore the direction of arcs and count the number of negative arcs in each semicycle (a closed semipath). In Figure 2, the sequence of arcs from P to X, on to O, and back to P constitute a semipath and a semicycle but not a path and a cycle, because the arcs do not point in the same direction. The semicycle is unbalanced because it contains an uneven number of negative arcs.

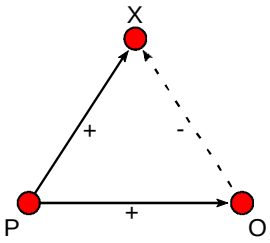


Figure 2 - P-O-X triple as a signed digraph.

- A **cycle** is a closed path.
- A **semicycle** is a closed semipath.
- A (semi)cycle is **balanced** if it does not contain an uneven number of negative arcs.

Fritz Heider was concerned with the feelings and perceptions of one person. Therefore, Figure 2 contains affections from person P to the other (O) and to the object or topic X. Even O's relation to X is measured from the perspective of P: it is P's idea about what O thinks of X, which does not necessarily correspond to O's real opinion. In social psychology, this phenomenon is called *attribution*. Of course, O may have positive or negative affect for P as well, and if X is a human being (or an animal) rather than a topic, X may also express affections for P and O.

Network analysts are interested in the feelings of all members of a group towards each other. This has lead to the notion of **structural balance** which

expects balance in the overall pattern of affect relations within a human group rather than in one person's affections and attributions.

Cartwright and Harary formulated exact conditions for a signed graph to be balanced. They noted that a balanced signed graph can be partitioned into two clusters such that all positive arcs are contained within the clusters and all negative arcs are located between clusters. You might say that a balanced network is extremely polarized because it consists of two factions and actors only have positive relations with members of their own faction whereas they have negative relations with members of the other faction. Clusters group people who like each other but who dislike members of the other cluster. It is easy to check this in Figure 3, which uses colors to identify the clusters (note that bi-directional arcs are replaced by edges).

In addition, they proved that a signed graph is balanced is all of its semicycles are balanced. Find one unbalanced semicycle and you know that the network is unbalanced.

- A **signed graph** is balanced if all of its (semi)cycles are balanced.
- A **signed graph** is balanced if it can be partitioned into two clusters such that all positive relations are contained within the clusters and all negative relations are situated between the clusters.

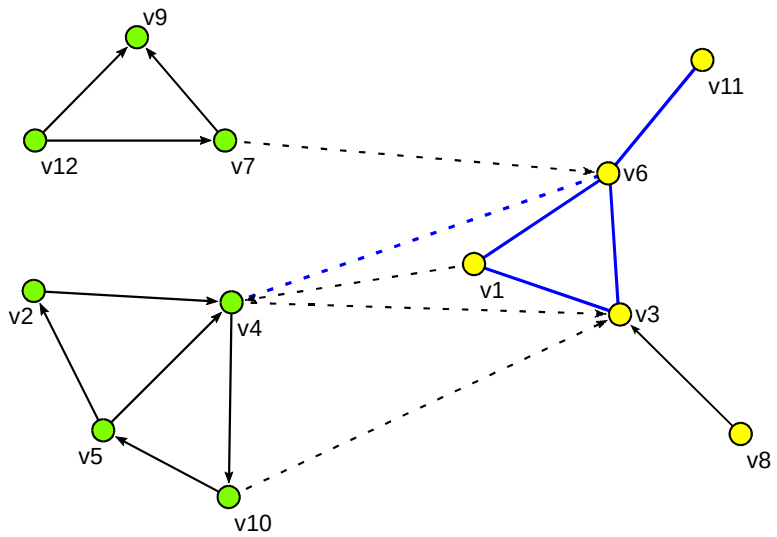


Figure 3 - A balanced network.

But why would human groups consist of two clusters or factions instead of three or more? In Figure 3, for instance, vertices v7, v9, and v12 could very well be a cluster on their own. In order to allow for three or more clusters, balance was generalized to **clusterability**. A signed network is clusterable if there is a partition satisfying the criterion that positive lines connect vertices within a cluster and negative lines are incident with vertices in different clusters, no matter the number of clusters. The network analyst Davis proved that a network is clusterable if it does not contain semicycles with exactly one negative arc.

Clearly, balance is a special case of clusterability because all balanced semicycles are clusterable.

- A cycle or semicycle is **clusterable** if it does not contain exactly one negative arc.
- A signed network is **clusterable** if it can be partitioned into clusters such that all positive relations are contained within clusters and all negative relations are situated between clusters.

In the course of time, balance theory has been generalized to models which incorporate hierarchy. We will present these models in Chapter 10. Some of them apply to unsigned networks but we will analyze signed relations only in the current chapter. In order to find subgroups in unsigned networks, we advise using the techniques for tracing cohesive subgroups presented in Chapter 3.

4.3 Example

In this chapter, we use a case which has been reanalyzed by network analysts many times, viz. the ethnographic study of community structure in a New England monastery by Samuel F. Sampson. The study describes several social relations among a group of 18 men (novices) who were preparing to join a monastic order. We will use the affect relations among the novices, which were collected by asking them to indicate whom they liked most and whom they liked least. The novices were asked for a first, second, and third choice on both questions.

The social relations were measured for several moments in time. The file `Sampson.net` contains the affect relations at four different moments. In the present section, however, we will focus on the affect relations between the novices at the fourth moment of time (T4), which was one week before four of them were expelled from the monastery. For the sake of illustration, we use their first choices only. The data are available in the file `Sampson_T4.net`.

Some novices had attended the minor seminary of ‘Cloisterville’ before they came to the monastery; they are identified as class one in the partition `Sampson_cloisterville.clu`. Based on his observations and analyses, Sampson divided the novices into four groups, which are represented by classes in the partition `Sampson_factions.clu`: Young Turks (class 1), Loyal Opposition (class 2), Outcasts (class 3), and an interstitial group (class 4). The Loyal Opposition consists of the novices who entered the monastery first. The Young Turks arrived later, in a period of change. They questioned practices in the monastery, which the members of the Loyal Opposition defended. Some novices did not take sides in this debate, so they are labeled ‘interstitial’. The Outcasts are novices who were not accepted in the group.

The Pajek project file `Sampson.paj` contains all networks and partitions.

4.4 Detecting structural balance and clusterability

Social networks are seldom perfectly balanced or clusterable. In some applications, researchers want to know whether a social network is more balanced or clusterable than we may expect by chance. If so, they conclude that the actors in the network adapt their relations to balance. In exploratory social network analysis, however, we are primarily interested in detecting balanced clusters, which represent cohesive subgroups within the network.

There are several ways to detect clusters in a signed network such that positive lines are within clusters and negative lines between clusters. Sometimes, clusters can be found by visual exploration. If we draw positive lines, which indicate attraction, as short as possible and negative lines, which signal repulsion, as long as possible, clusters of positive relations are clearly visible in a sociogram. In Figure 4, which is drawn in this manner, we can see three clusters in the network of novices. Because there are three clusters, the network is clusterable rather than balanced.

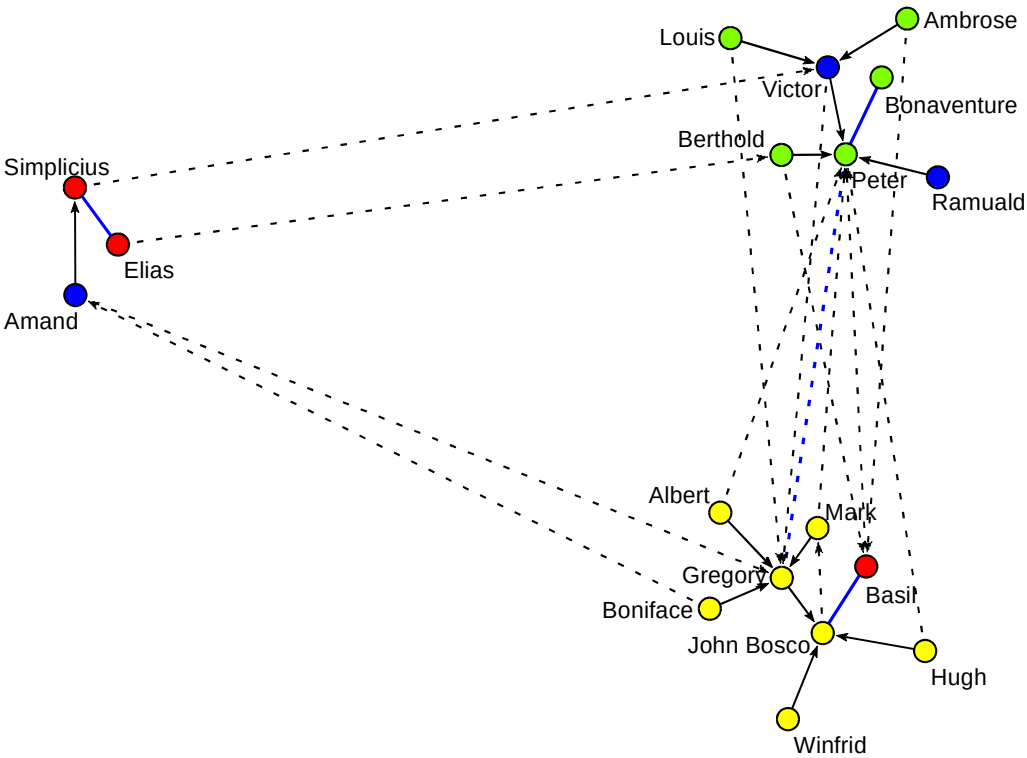


Figure 4 - First positive and negative choices between novices at time four (T4).

Because the network is highly clusterable and not very dense, we can visually check that all positive arcs are situated within clusters and almost all negative arcs are directed from one cluster to another. The only negative arc within a cluster points from John Bosco to Mark at the bottom of the sociogram. Note that the triple John Bosco, Mark, and Gregory contains exactly one negative arc, so it is unclusterable and it will yield problems in any clustering which we may attempt.

In Figure 4, vertex colors indicate the factions which Sampson delineated: Young Turks (yellow), Loyal Opposition (green), Outcasts (red), and the

interstitial group (blue). The social cleavage between the Young Turks (yellow) and the Loyal Opposition (green) is evident, but the Outcasts are not clustered perfectly. Ramuald and Victor are clustered with the Loyal Opposition and they probably felt somewhat related to them because they all (except Louis) came from 'Cloisterville'.

If sociograms are not as orderly as Figure 4, we must use computational techniques to find the clustering which fits balance or clusterability best. In exploratory network analysis, a good strategy is to try many clusterings and select the one containing the lowest number of forbidden lines: positive lines between clusters or negative lines within a cluster. The number of forbidden lines is an error score which measures the degree of balance or clusterability in a network: more errors mean less balance or clusterability.

In Figure 4, there is just one forbidden line if we partition the novices into three clusters, namely, the negative arc from John Bosco to Mark in the bottom right cluster. It is up to the researcher to decide whether the degree of balance or clusterability is acceptable. Criteria cannot be specified without the use of estimation techniques, which fall outside the scope of this book, because the acceptability of an error score depends on the size and density of a network. The error score allows us to pick the best fitting clustering, but it does not say whether it is good enough.

The approach of rearranging vertices into clusters over and over again and selecting the best solution, is an **optimization technique**, which has three features that are worth noting. First, an optimization technique may find several solutions or partitions which fit equally well. It is up to the researcher to select one or present them all.

Second, it is possible that this technique does not find the best fitting clustering, although this is expected to happen only in exceptional cases. Nevertheless, there is no guarantee that there is not a better solution, unless, of course, you find a clustering which fits perfectly. We advise to repeat the procedure many times and to inspect the results visually in order to see whether you can find a better solution.

Third, starting options may yield different results, for instance, the procedure finds another solution if it is told to look for two clusters instead of three or four. It is usually possible to estimate the approximate number of clusters from an energized sociogram, but it is hard to tell the exact number of clusters which will yield the lowest error score. Therefore, it is important to repeat the optimization technique with different numbers of clusters.

In addition, the user may attribute different weights or penalties to forbidden positive and negative arcs. For instance, researchers have noted that negative arcs within a cluster are tolerated less than positive arcs between clusters, so we can raise the penalty on negative arcs within clusters. In the network of affect relations between novices depicted in Figure 4, this would mean that John Bosco's negative feelings for Mark are more important than Gregory's positive affection for John Bosco. Hence, the optimization technique will split the bottom

cluster between John Bosco and Gregory. Different weights may produce different results.

Application

[Draw screen] Options >Values of Lines>Similarities

A sociogram which minimizes the length of positive lines and maximizes the length of negative lines can be made in two steps. First, select the option *Similarities* in the *Options>Values of Lines* submenu of the Draw screen. This option tells the energy procedures that line values indicate similarity or attraction: the higher a line value, the closer two vertices should be drawn. Negative line values mean that vertices are dissimilar and must be drawn far apart. In Pajek, signs of lines are represented as line values 1 and -1, so positive arcs are short and negative arcs are long in an energized drawing. Note that this option remains effective until another option is selected. Second, apply an energy procedure to the sociogram.

The command *Balance*, which searches an optimal clustering in a signed network, is located in the *Operations* menu because it requires two different data objects: a network and a partition. The network contains the vertices and relations that must be clustered and the partition specifies the number of clusters and the initial clustering which the computer tries to improve.

Partition >Create Random Partition

If you do not have a partition with a meaningful initial clustering, you can easily make a random partition with the *Create Random Partition* command in the *Partition* menu. Just specify the number of vertices, which should be equal to the number of vertices in the network, and the number of clusters you want to detect in the network.

Operations>Balance

The *Balance* command asks how many times it must try to find an optimal clustering. In each repetition, it starts with a new random partition. If a starting partition fits quite well, the optimization technique will not find better solutions because all changes will increase the error score initially. This could happen, for instance, with a starting clustering based on visual inspection of the energized sociogram. With several random starting partitions, the procedure is unlikely to miss a good clustering which differs a lot from your expectations, although this is not guaranteed. In a small network, 100 repetitions is a reasonable choice.

Next, you have to specify the error weight of a forbidden negative arc, that is, a negative arc within a positive cluster. This weight is called alpha and it is .5 by default. The error weight for an erroneous positive arc is equal to one minus alpha, so negative and positive arcs are treated equally by default. If you want to penalize a forbidden negative arc more than an erroneous positive arc, raise alpha in the dialog box, for instance, to .75. In consequence, a forbidden positive arc is weighted by .25, which is a third of the weight attached to an out-of-place negative arc.

Figure 5 shows the results for the novices network. We used Sampson's classification as a starting partition with four clusters (`Sampson_factions.clu`) and it was instructed to weigh positive and negative errors equally. First, the listing displays the error score and the erroneous arcs in

the initial clustering. There are quite a lot of errors, which are identified by vertex numbers in the listing of lines. The procedure finds two solutions with exactly one ‘forbidden’ arc. In the first clustering, a positive arc connects vertices 7 (Mark) and 2 (Gregory), which are apparently members of different clusters. In the second clustering, the positive arc from vertex 2 (Gregory) to vertex 1 (John Bosco) is a problem. As expected, the unclusterable triple John Bosco - Mark - Gregory causes these problems. Nevertheless, the clustering is nearly perfect, so we may conclude that the network is clusterable. In order to know whether it is balanced as well, we must repeat the procedure with a starting partition containing two clusters.

```
-----------------------------------------------------------------------
Partitioning Signed Graphs
-----------------------------------------------------------------------
 Working...
Number of classes: 4, alpha: 0.500
-------  Starting partition -------
Errors:       4.00       Lines
-----------------------------------
             -1.00 :      1.7
              1.00 :      1.3
              1.00 :      3.1
              1.00 :      8.4
              1.00 :      9.8
              1.00 :     10.4
              1.00 :     11.8
              1.00 :     13.18
-----------------------------------
-------     Improvements    -------
     1:       0.50
-------   Final partition  1-------
Errors:       0.50       Lines
-----------------------------------
              1.00 :      7.2
-----------------------------------
-------   Final partition  2-------
Errors:       0.50       Lines
-----------------------------------
              1.00 :      2.1
-----------------------------------
 2 solutions with 0.50 errors found.
 Time spent:  0:00
```

Figure 5 - Output listing of a *Balance* command.

All optimal solutions are saved as partitions in the *Partitions* drop list. Drawing the network with these partitions, we can see that clusters at the left and at the top are correctly identified. The cluster at the bottom is split in two ways: either Mark is considered to be a cluster on his own, or he is grouped with Albert, Boniface, and Gregory, who are separated from John Bosco, Basil, Hugh, and Winfrid. The last clustering is most likely if negative arcs within a cluster are considered slightly more problematic than positive arcs between clusters. For instance, try *Balance* with alpha set to .6 .

Let us conclude this section with a warning. The *Balance* command triggers a procedure which is very time-consuming, so it should not be applied to networks with more than a hundred vertices unless you do not need your computer for some hours or days. In Pajek, commands which should only be applied to small networks are marked by a * in the menu.

4.5 Development in time

Balance theory expects a tendency towards balance. In the course of time, affect relations within a human group are hypothesized to become more balanced or clusterable. This raises the question of how to analyze the evolution of social networks. In this section, we will discuss the simplest way to analyze longitudinal networks, namely by comparing network structure at different points in time.

Sampson measured affect relations at the monastery at five moments. At the time of the first measurement (T1), the group consisted mainly of novices who soon left the monastery to study elsewhere. Therefore, we will not use this network and it is not included in the file `Sampson.net`. The second measurement (T2) concerns the period just after the arrival of a number of newcomers. The third measurement (T3) shows affect relations around the time when one of the newcomers (Brother Gregory) organized a meeting to discuss the situation in the monastery. This meeting fuelled a process of polarization, which led to the expulsion of four novices (among them Brother Gregory) one week after the fourth measurement (T4). The expulsion triggered the voluntary departure of many novices in the next few weeks. At the time of the fifth measurement (T5), no more than seven of the eighteen novices were still living in the monastery, so this network is difficult to compare to the previous networks. We will analyze the networks at T2, T3, and T4 only.

```
*Vertices     18
      1 "John Bosco"      0.1000    0.5000    0.5000 [2-4]
      2 "Gregory"         0.1241    0.3632    0.5000 [2-4]
      3 "Basil"           0.1936    0.2429    0.5000 [2-4]
      4 "Peter"           0.3000    0.1536    0.5000 [1-5]
      5 "Bonaventure"     0.4305    0.1061    0.5000 [1-*]
      6 "Berthold"        0.5695    0.1061    0.5000 [1-*]
      7 "Mark"            0.7000    0.1536    0.5000 [1-4]
      8 "Victor"          0.8064    0.2429    0.5000 [1-4]
      9 "Ambrose"         0.8759    0.3632    0.5000 [1-*]
     10 "Ramuald"         0.9000    0.5000    0.5000 [2-5]
     11 "Louis"           0.8759    0.6368    0.5000 [2-*]
     12 "Winfrid"         0.8064    0.7571    0.5000 [2-5]
     13 "Amand"           0.7000    0.8464    0.5000 [2-4]
     14 "Hugh"            0.5695    0.8939    0.5000 [2-4]
     15 "Boniface"        0.4305    0.8939    0.5000 [2-4]
     16 "Albert"          0.3000    0.8464    0.5000 [2-4]
     17 "Elias"           0.1936    0.7571    0.5000 [2-4]
     18 "Simplicius"      0.1241    0.6368    0.5000 [2-4]
*Arcs
      1      2       2  [3]
      1      2      -2  [4]
      1      3       2  [2]
      1      3       3  [4]
      1      4      -2  [3]
      1      5       1  [3]
      1      5       3  [2]
      1      6      -2  [2]
      1      7      -1  [2]
      1      7      -3  [4]
      1      8       3  [3]
      1     10      -1  [4]
      1     10      -3  [2,3]
```

Figure 6 - Partial listing of Sampson.net.

Application

Pajek has special facilities for longitudinal networks. Figure 6 shows part of the network file `Sampson.net`. By now, the structure of the lists of vertices and arcs will be familiar to you, so you can focus on the time indicators in square brackets which are added to each vertex and arc. For instance, Brother John Bosco was at the monastery from time 2 up to and including time 4. He left before time 5. Brother Bonaventure arrived at the monastery before the first measurement and he stayed after the last measurement. The star (*) indicates infinity. At time three, the arc from vertex 1 (John Bosco) to vertex 2 (Gregory) has a value of 2, indicating a positive second choice. At time four, however, it has turned into a second negative choice (line value -2). At times two and three, John Bosco chooses Ramuald (vertex 10) as the person he likes least (line value -3). Note that time is always represented by positive integers.

Net>Transform >Generate in Time

The time notation can be used to split the longitudinal network into separate networks for different moments or periods. The submenu *Net>Transform>Generate in Time* offers the user several commands for generating a series of cross-sectional networks. First of all, you can choose to obtain a network for each period requested (option *All*) or to produce a network only if it differs from the previous one (option *Only Different*). The latter command is useful if a network does not change much over time. Whichever command you choose, you will have to specify the first and last time point which you want to analyze as well as the time interval (step) between successive networks. In our example, we start at time two, stop at time 4, and we want a network for each moment in between, so we choose step value one. Step values must be positive integers. With a step value of two, the command creates a network for the first, third, fifth moment, et cetera.

Previous

Next

Options>Previous/Next >Apply to

After generating networks for separate moments, you can easily switch from one moment to another with the commands *Previous* and *Next*, provided that *Network* is the only option selected in the *Options>Previous/Next>Apply to* submenu in the Draw screen . It is sensible to inspect all generated networks to check whether the results are as intended. Errors in time indicators may have serious consequences, for instance, if Brother John Bosco is not present at time three - by mistake, the indicator reads [2,4] instead of [2-4] - all arcs to and from him will be deleted from the network at time three. Note also that serial numbers of vertices change in generated networks when vertices disappear from the network (as at time five) or when new vertices are added, since numbers of vertices must always range from 1 to the number of vertices in Pajek.

Table 1 - Error score with all choices at different moments (α = .5).

	Time Points		
Number of clusters	T2	T3	T4
2 (balance)	21.5	16.0	12.5
3	17.5	11.0	10.5
4	19.0	13.5	12.5
5	20.5	16.0	15.0

Having generated networks for times two, three, and four, we can analyze the degree of balance or clusterability in each period. Table 1 shows the error scores associated with the optimal clustering for several numbers of clusters and different time points. Note that the network now consists of all three positive and negative choices at times T2, T3, and T4. First choices count three times (line values are 3 and -3 for the first positive and negative choice respectively), second choices count double (values 2 or -2), and third choices count once (1 or -1). The error score is computed from these line values, which explains why it is quite large compared to error scores in the previous section: a forbidden first choice contributes .5 (alpha) times 3 instead of .5 times 1 to the error score.

Table 1 helps us to draw a conclusion on the evolution of balance and clusterability in the network of novices. Each clustering fits better in the course of time and the partition with three clusters fits best at any moment, so we may conclude that there is a tendency towards clusterability rather than balance. This is probably due to a process of polarization, which ends when several novices leave the monastery. Instead of a continuing trend towards balance or clusterability, human groups probably experience limited periods of polarization, which are reflected in increasingly balanced or clusterable patterns of affective relations.

4.6 Summary

In this chapter, we discuss cohesive subgroups in signed networks, that is, in networks with positive and negative relations. If relations represent affections, people who like each other tend to huddle together, whereas negative sentiments exist predominantly between groups. This principle stems from balance theory and has been generalized to groups with three or more clusters (clusterability). Balance and clusterability occur in times of polarization between opposing factions.

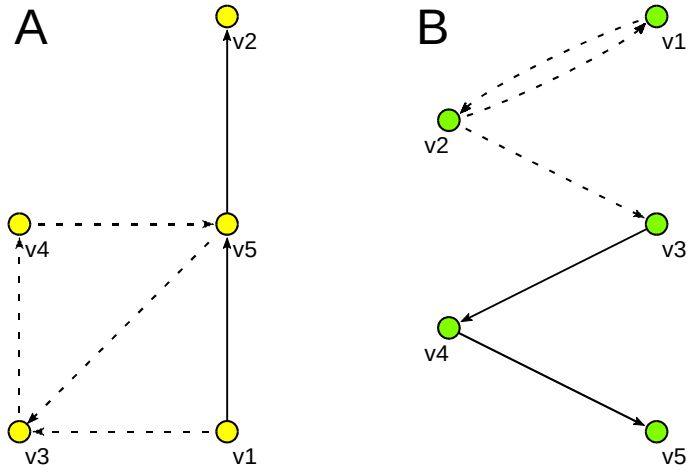
We may determine whether affective relations are patterned according to balance theory by searching for a partition or clustering which satisfies the principle that positive lines are found within clusters and negative lines between clusters. If a signed network can be partitioned into two clusters according to this principle, the network is balanced. A partition with three or more clusters is a clusterable network.

The weighted number of lines which do not conform to the balance theoretic principle, namely, negative lines within a cluster and positive lines between clusters, indicates the degree of balance or clusterability in a network. This is called the error score of the best fitting clustering. For one network, we compare error scores for different numbers of clusters to find the optimal clustering. If the error score is acceptable, the clusters represent cohesive groups. In addition, we may compare error scores of a network at different moments to check whether group structure displays a tendency towards balance which is predicted by balance theory.

4.7 Exercises

1 Which of the following statements is correct?
   a a signed graph is balanced if it is clusterable into two clusters
   b a signed graph is balanced if it is not clusterable
   c a signed graph is balanced if its vertices can be partitioned into two groups
   d a signed graph cannot be balanced because it is undirected

2 Have a look at the following two networks and check which of the following statements is correct.



   a A and B are balanced
   b A is balanced and B is clusterable
   c A is clusterable and B is balanced
   d neither A nor B is balanced

3 How many cycles and semicycles does network A of Exercise 2 contain?
   a no cycles and two semicycles
   b one cycle and one semicycle
   c one cycle and two semicycles
   d two cycles and one semicycle

4 Which of the following statements about the sequence of lines v2-v3-v4-v5 in network B (Exercise 2) is correct?
   a it is neither a semiwalk nor a semipath
   b it is a semiwalk and a semipath
   c it is a walk and a semipath
   d it is a walk and a path

5 An analysis of 58 alliance and 58 antagonistic relations among 16 tribes in New Guinea yields error scores reported in the table below (alpha is .5). Which conclusion do you support? Please, state your reasons.

	number of clusters			
	2	3	4	5
error score	7.0	2.0	4.0	6.0

a the tribal network is balanced
b the tribal network is clusterable
c the tribal network is neither balanced nor clusterable
d it is impossible to draw a conclusion from these results

6 In the best fitting clustering presented in exercise 6, how many arcs violate the balance theoretic principle?
a 2 arcs
b 4 arcs
c 6 arcs
d 7 arcs

4.8 Assignment

In 1943, Leslie D. Zeleny administered a sociometric test to 48 cadet pilots at an US Army Air Forces flying school. Cadets were trained to fly a two-seated aircraft, taking turns in flying and aerial observing. Cadets were assigned to instruction groups ranging in size from 5 to 7 at random, so they had little or no control over who their flying partners would be. The sociometric test was used to improve the composition of instruction groups. Zeleny asked each cadet to name the members of his flight group with whom he would like to fly as well as those with whom he would not like to fly.

The data are available in the project file `Flying_teams.paj` which contains a network (`Flying_teams.net`) and the original (alphabetical) instruction groups (`Flying_teams.clu`).

Which instruction groups would you advice in order to reduce the risk of cadets flying with partners they do not want? Try to find groups of 5 to 10 cadets.

4.9 Further reading

- The example was taken from S.F. Sampson, *A Novitiate in a Period of Change. An Experimental and Case Study of Social Relationships* (PhD thesis Cornell University, 1968). An analysis of the data using the method presented in this chapter can be found in P. Doreian and A. Mrvar, ‘A partitioning approach to structural balance’ (in *Social Networks*, 18 (1996), p. 149-168).
- The New Guinea data were reported in K. Read, ‘Culture of the central highlands, New Guinea’, in the journal *Southwestern Journal of Anthropology* (vol. 10 (1954), p. 1-43) and reanalyzed in P. Hage and F. Harary, *Structural models in anthropology* (Cambridge/New York: Cambridge University Press, 1983). The data are also distributed with UCINET network analysis software.
- The flying teams data are taken from J. L . Moreno (et al.), *The Sociometry Reader* (Glencoe (Illinois): The Free Press, 1960, p. 534-547).
- For an overview of balance theory in social network analysis, consult Chapter 6 in S. Wasserman and K. Faust, *Social Network Analysis: Methods and*

Applications (Cambridge: Cambridge University Press, 1994). They also give references for Davis' proof.

- F. Harary, R.Z. Norman, and D. Cartwright, *Structural Models : An Introduction to the Theory of Directed Graphs* (New York: Wiley, 1965) contains the mathematics discussed here. The lucidity of this book makes it a pleasure to read. In addition, *Discrete Mathematical Models* by F.S. Robberts (New York: Prentice Hall, 1976) is a good book to further your understanding of several mathematical models including balance.

4.10 Answers

1 Answer a is correct. Balance is a special case of clusterability, because a network is balanced if it can be partitioned into two clusters in such a way that all positive lines are within clusters and all negative lines are between clusters.

2 Answer c is correct. Semicycles v1-v3-v4-v5 and v1-v3-v5 contain three negative arcs, so network A is not balanced. Cycle v3-v4-v5 has two negative arcs. There is no (semi)cycle with exactly one negative arc, so network A is clusterable. Network B contains just one (semi)cycle, namely v1-v2, which contains two negative arcs. Hence, network B is balanced. Network A can be partitioned into three clusters according to the principle of positive arcs within clusters and negative arcs between clusters: v1, v2, and v5 are a cluster, v3 and v4 are separate clusters. Likewise, network B can be divided into two clusters: v2 versus the rest.

3 Answer c is correct, see the answer to exercise 3.

4 Answer d is correct. The sequence of lines is a walk because the lines are adjacent and they point in the same direction: the head of one arc is the tail of the next arc. This walk is a path because all vertices are distinct, meaning that no vertex occurs more than once in the walk.

5 Answer b is the most likely conclusion. Clearly, a partition with three clusters fits better than a partition with two clusters. Therefore, the network is clusterable rather than balanced. However, you might decide that even an error score of 2.0 is too high and conclude that the network is neither balanced nor clusterable. This would be very strict, though, because four erroneous arcs (see exercise 7) on a total of 116 arcs is quite low.

6 Answer b is correct. Alpha is .5, which means that every 'forbidden' arc - positive or negative - contributes .5 to the error score. Therefore, four arcs in the wrong places yield a total error score of 2.0, which is the minimal error score in the table.