

# Balance in Stochastic Signed Graphs

Ove Frank

*University of Lund\**

Frank Harary

*University of Michigan\*\**

*Signed graphs provide models for investigating balance in connection with various kinds of social relations. Since empirical social networks always involve uncertainty because of errors due to measurement, imperfect observation or sampling, it is desirable to incorporate uncertainty into signed graph models. We introduce a stochastic signed graph and investigate the properties of some indices of balance involving triads. In particular we consider the balance properties of a graph which is randomly signed and of one which has been randomly sampled from a large population graph.*

## Introduction

The notion of balance in social networks was introduced in the classical paper by Heider (1946) and in his book (Heider, 1958) and later was the subject of various theories and discussions of social behavior and social relations, as for example in Cartwright and Harary (1956), Abelson and Rosenberg (1958) and Zajonc (1968).

A mathematical framework for the investigation of balance was developed by Harary (1953) who introduced the concept of a signed graph. Balance in signed graphs was further explored in a series of papers by Harary (1955, 1957, 1959, 1960) and Cartwright and Harary (1968, 1970) which discuss various aspects of the measurement of balance, conditions for balance and consequences of balance in graph-theoretic terms.

The deterministic signed graph approach to the modelling of balance implies that the practical issues of measurement errors and other sources of uncertainty in empirical investigations have to be ignored. It is possible to include the nondeterministic influence of uncertainty in the models by intro-

---

\*Department of Statistics, University of Lund, S-220 07 Lund, Sweden.

\*\*Department of Mathematics, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.

This work has been partially supported by the Swedish Council for Social Science Research and the National Science Foundation, Sociology Division, U.S.A.

ducing stochastic signed graphs. Stochastic graphs have been used by Frank (1979a, b), Holland and Leinhardt (1979) and others as a tool for describing social networks which involve uncertainty of different kinds.

A particular kind of uncertainty appears if the empirical network under study has been selected by simple random sampling from a larger unknown population network. Frank (1971, 1977a, b, c, 1978a, b) has considered various inference problems based on different kinds of available sample information.

Another kind of uncertainty which may be due to measurement or observational errors can be investigated by using stochastic graphs with simple disturbances of the edge occurrences. Such models have been used by Frank (1978c, 1979c, d) for analysing clustering and transitivity properties in networks.

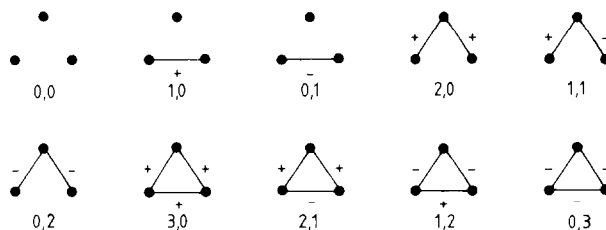
Here we will apply a simple stochastic signed graph model in order to investigate the concept of balance without ignoring the uncertainty which is present in empirical networks. Section 2 gives some basic terminology on balance and Section 3 introduces a stochastic signed graph model. The properties of some convenient indices of balance are investigated in Section 4, and expected values and variances are derived for the numbers of positive and negative 3-cycles. The results are used in Section 5 to find approximate formulae pertaining to two special cases of some general interest, namely the case of a graph which has its edges signed at random and the case of a graph which has been selected by sampling from a population graph.

## Balance

A graph  $G$  is signed if each edge is given either a positive or negative sign. In a signed graph  $G$  there can be ten different types of induced subgraphs of order 3 as shown in Fig. 1. These induced subgraphs are called the triads of  $G$ . For each triad let  $r$  and  $s$  be the number of positive and negative edges in it. Then we denote the frequencies of the different triads in  $G$  by  $t_{rs}(G)$  for all possible combinations of  $r$  and  $s$ . These frequencies are called the triad counts.

The sign of a cycle in a signed graph  $G$  is the product of its edge signs. A signed graph  $G$  is balanced if all its cycles are positive.

Figure 1. *Signed triads with (r,s)-labels.*



The degree of balance in a signed graph  $G$  can be measured in various ways, and conveniently by examining the triads and particularly the 3-cycles. Let  $c^+(G)$  and  $c^-(G)$  be the numbers of positive and negative 3-cycles in  $G$ , *i.e.*,

$$c^+ = t_{30} + t_{12}, \quad c^- = t_{21} + t_{03}.$$

One index of balance is given (Harary, 1959) by the proportion of positive (balanced) 3-cycles which is

$$B_1 = c^+ / (c^+ + c^-)$$

if there are any 3-cycles, and which is defined as 1 otherwise. Another index of balance is given by the proportion of balanced triads, *i.e.*,

$$B_2 = 1 - c^- / \binom{N}{3},$$

where  $N$  is the order of the graph  $G$ . This proportion is seen to be at least as large as the proportion of balanced 3-cycles with equality if and only if  $c^- = 0$  or  $c^+ + c^- = \binom{N}{3}$ . Both the proportion of balanced 3-cycles and the proportion of balanced triads range between 0 and 1, and they are equal to 1 if and only if there are no negative 3-cycles.

Our object is to investigate balance in terms of triads for stochastic signed graphs, and it will then prove advantageous not to use the relative but the absolute numbers of balanced 3-cycles and balanced triads, *i.e.*,  $c^+$  and  $\binom{N}{3} - c^-$ . We will use the expected values and variances of these numbers in order to find approximate formulae for the expected values and variances of  $B_1$  and  $B_2$ .

### A stochastic graph model

Let  $G$  be a graph of order  $N$  with node set  $V = \{1, \dots, N\}$  and adjacency indicators

$$a_{ij} = \begin{cases} 1 & \text{if nodes } i \text{ and } j \text{ are adjacent,} \\ 0 & \text{otherwise.} \end{cases}$$

We have  $a_{ii} = 0$  and  $a_{ij} = a_{ji}$  for all  $i$  and  $j$  in  $V$ . We will assume that  $G$  is a stochastic graph having  $\binom{N}{2}$  independent Bernoulli ( $\alpha$ )-distributed  $a_{ij}$  for  $i < j$ :

$$a_{ij} = \begin{cases} 1 & \text{with probability } \alpha, \\ 0 & \text{with probability } 1 - \alpha. \end{cases}$$

The graph  $G$  will further be assumed to be stochastically signed, *i.e.*, there are  $\binom{N}{2}$  independent stochastic sign variables

$$s_{ij} = \begin{cases} 1 & \text{with probability } p, \\ -1 & \text{with probability } q = 1 - p, \end{cases}$$

for  $i < j$ , which yield the signs of all node pairs. The sign variables  $s_{ij}$  are stochastically independent of the adjacency indicators  $a_{ij}$ . Let  $s_{ii} = 0$  and  $s_{ij} = s_{ji}$ . We now introduce the signed adjacency variables  $x_{ij} = s_{ij}a_{ij}$  for all  $i$

and  $j$  in  $V$ . These variables  $x_{ij}$  which indicate the signs of the edges in  $G$  are independent for  $i < j$  and their distribution is given by

$$x_{ij} = \begin{cases} 1 & \text{with probability } p\alpha, \\ -1 & \text{with probability } q\alpha, \\ 0 & \text{with probability } 1 - \alpha. \end{cases}$$

In order to be able to consider subgraphs of  $G$  induced by a node sample  $S \subseteq V$ , we introduce the node selection variables

$$y_i = \begin{cases} 1 & \text{if node } i \text{ belongs to the sample } S, \\ 0 & \text{otherwise.} \end{cases}$$

The subgraph of  $G$  induced by  $S$  will be denoted by  $G(S)$ , and it is determined by the signed adjacency variables  $z_{ij} = x_{ij}y_iy_j$ .

If the sample consists of  $n$  nodes selected by simple random sampling from  $V$ , it is well known that the sampling procedure can be approximated by a Bernoulli ( $\beta$ )-sampling scheme with selection probability  $\beta = n/N$ . Thus, for each node we decide independently and with a common probability  $\beta > 0$  whether or not the node should be included in the sample; see for instance Frank (1971). We will use this Bernoulli-sampling scheme since it leads to simpler formulae which for our purposes are satisfactory approximations for the more cumbersome formulae pertaining to simple random sampling. Thus we assume that the  $y_i$  are independent for  $i \in V$  and

$$y_i = \begin{cases} 1 & \text{with probability } \beta, \\ 0 & \text{with probability } 1 - \beta. \end{cases}$$

Moreover, the  $y_i$  are of course independent of the  $a_{ij}$  and  $s_{ij}$ . It follows that the signed adjacency variables of  $G(S)$  satisfy

$$z_{ij} = \begin{cases} 1 & \text{with probability } p\alpha\beta^2, \\ -1 & \text{with probability } q\alpha\beta^2, \\ 0 & \text{with probability } 1 - \alpha\beta^2, \end{cases}$$

and they are not independent but have a simple dependence structure given by the product formula

$$z_{ij} = s_{ij}a_{ij}y_iy_j$$

involving independent sign, adjacency and selection variables.

### Triads and balance

In this section we will consider the subgraph  $G(S)$  induced by a sample  $S$ , and it is tacitly understood that  $t_{rs}$  refers to  $t_{rs}(G(S))$ ,  $c^+$  to  $c^+(G(S))$ , etc.

Using the general results on stochastic graph triads given by Holland and Leinhardt (1975) and Frank (1979b), we can deduce the expected values and the variances and covariances of the triad counts  $t_{rs}$ . We will be content here with reporting only the expected values given in the following theorem, the proof of which is straightforward and omitted.

*Theorem 1.* The graph  $G(S)$  has the expected triad counts

$$Et_{rs} = \binom{N}{3} \binom{3}{r+s} \binom{r+s}{r} p^r q^s \alpha^{r+s} (1 - \alpha)^{3-r-s} \beta^3$$

for all possible combinations of  $r$  and  $s$  according to Figure 1.

From Theorem 1 we find readily the expected numbers of positive and negative 3-cycles. The next theorem gives the expected value and the variance of the number of positive 3-cycles. We note that the corresponding result for negative 3-cycles is obtained by interchanging  $p$  and  $q$ .

*Theorem 2.* For the number of positive 3-cycles in the graph  $G(S)$ , the expected value is

$$Ec^+ = \binom{N}{3} P,$$

and the variance is

$$\text{Var } c^+ = \binom{N}{3} P(1 - P) + 12 \binom{N}{4} (Q - P^2) + 30 \binom{N}{5} P^2 (1 - \beta) / \beta,$$

where

$$P = (p^3 + 3pq^2) \alpha^3 \beta^3,$$

$$Q = (p^5 + 2p^3q^2 + 4p^2q^3 + pq^4) \alpha^5 \beta^4.$$

*Proof.* If we introduce

$$z_{ijk} = \begin{cases} 1 & \text{if } z_{ij}z_{jk}z_{ki} > 0, \\ 0 & \text{otherwise,} \end{cases}$$

we obtain

$$c^+ = \frac{1}{6} \sum_{V^{(3)}} z_{ijk}.$$

where  $V^{(3)}$  is the set of all ordered triples of distinct nodes in  $V$  and the sum is over  $(i, j, k) \in V^{(3)}$ . According to our stochastic model the  $z_{ijk}$  are independent for triples with no common nodes. They have a common expected value  $Ez_{ijk} = P$ , where  $P$  is readily seen to be equal to the expression given in the theorem. By examining the eight ways (see Fig. 2) to combine two positive 3-cycles with exactly two nodes in common, we find that  $Ez_{ijk}z_{jkl} = Q$  with  $Q$  as given in the theorem. Moreover, by examining the sixteen ways (see Fig. 3) to combine two positive 3-cycles with only one node in common we find that

$$\begin{aligned} Ez_{ijk}z_{klm} &= (p^6 + 6p^4q^2 + 9p^2q^4) \alpha^6 \beta^5 \\ &= (p^3 + 3pq^2)^2 \alpha^6 \beta^5 = P^2 / \beta. \end{aligned}$$

According to Frank (1979b, Theorem 3) we then find the values  $Ec^+$  and  $\text{Var } c^+$  given in the theorem.

Figure 2. *Combinations of two signed triads with two nodes in common.*

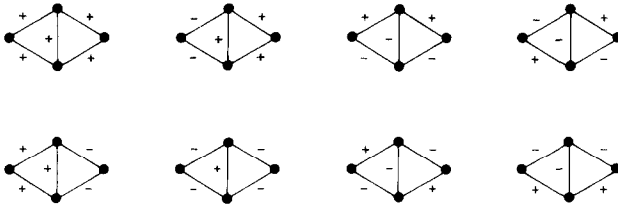
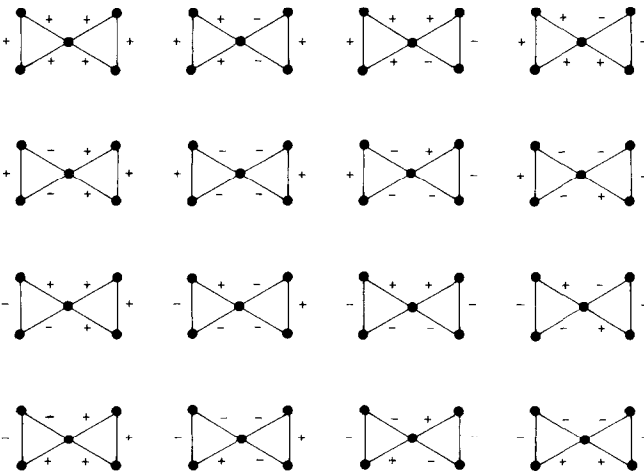


Figure 3. *Combinations of two signed triads with one node in common.*



### Applications

We first apply Theorem 2 to find the likely ranges of the indices of balance  $c^+$ ,  $B_1$  and  $B_2$  for a randomly signed graph. For this purpose we will use the deviation interval given by the expected value plus and minus the standard deviation.

Consider an empirical signed graph  $G$  having  $N$  nodes and  $R$  edges of which  $R^+$  are positive and  $R^-$  negative. We can consider  $N, R^+$  and  $R^-$  as fixed numbers and compare the empirical outcome of  $c^+$  with its likely range if the edges were signed at random. If we use our stochastic model with

$$p = R^+/R, \alpha = R / \binom{N}{2}, \beta = 1,$$

we find from Theorem 2 that  $c^+$  has the deviation interval

$$\binom{N}{3}P \pm \sqrt{\binom{N}{3}P(1-P) + 12\binom{N}{4}(Q - P^2)},$$

where

$$P = (p^3 + 3pq^2)\alpha^3,$$

$$Q = (p^5 + 2p^3q^2 + 4p^2q^3 + pq^4)\alpha^5.$$

If we consider a large graph ( $N$  large) and approximate  $c = c^+ + c^-$  by its expected value  $\binom{N}{3}\alpha^3$ , we find that the balance index  $B_1$  has the approximate deviation interval

$$p^3 + 3pq^2 \pm \frac{3}{N} \sqrt{\frac{2}{\alpha} (p^5 + 2p^3q^2 + 4p^2q^3 + pq^4) - 2(p^3 + 3pq^2)^2}$$

and  $B_2$  has the approximate deviation interval

$$1 - \alpha^3(3p^2q + q^3) \pm \frac{3\alpha^3}{N} \sqrt{\frac{2}{\alpha} (p^4q + 4p^3q^2 + 2p^2q^3 + q^5) - 2(3p^2q + q^3)^2}.$$

If we consider a complete graph with  $R = \binom{N}{2}$ , then we can use the model specification

$$p = 1/2, \alpha = 1, \beta = 1,$$

which yields the deviation interval for  $c^+$  as

$$\frac{1}{2} \binom{N}{3} \pm \frac{1}{2} \sqrt{\binom{N}{3}}$$

and the approximate deviation intervals for both  $B_1$  and  $B_2$  as

$$\frac{1}{2} \pm \frac{1}{N} \sqrt{\frac{3}{2N}}.$$

Let us now turn to the situation where the empirical graph  $G(S)$  is the subgraph induced by a simple random sample of  $n$  nodes from a large population graph  $G$  of known order  $N$ . Let  $r^+$  and  $r^-$  denote the numbers of positive and negative edges in  $G(S)$ , and specify the stochastic model by

$$p = r^+/r, \alpha = r/\binom{n}{2}, \beta = n/N < 1,$$

where  $r = r^+ + r^-$ . It follows that the approximate deviation intervals are for  $c^+$ ,

$$\frac{4r^3}{3n^3}(p^3 + 3pq^2) \pm \frac{4r^3}{n^3}(p^3 + 3pq^2) \sqrt{\frac{N-n}{Nn}},$$

for  $B_1$ ,

$$p^3 + 3pq^2 \pm 3(p^3 + 3pq^2) \sqrt{\frac{N-n}{Nn}},$$

and for  $B_2$ ,

$$1 - \frac{8r^3}{n^3(n-1)^3}(3p^2q + q^3) \pm \frac{24r^3}{n^3(n-1)^3}(3p^2q + q^3) \sqrt{\frac{N-n}{Nn}}.$$

If the population graph is assumed to have the same number of positive and negative edges, then we can use the model specification

$$p = 1/2, \alpha = r/\binom{n}{2}, \beta = n/N < 1,$$

which yields for  $c^+$  the deviation interval

$$\frac{1}{2} \binom{N}{3} \alpha^3 \beta^3 \pm \frac{1}{2} \sqrt{\binom{N}{3} \alpha^3 \beta^3 (2 - \alpha^3 \beta^3) + 12 \binom{N}{4} \alpha^5 \beta^4 (1 - \alpha \beta^2) + 30 \binom{N}{5} \alpha^6 \beta^5 (1 - \beta)},$$

for  $B_1$  the approximate deviation interval

$$\frac{1}{2} \pm \frac{3}{2} \sqrt{\frac{N-n}{Nn}},$$

and for  $B_2$  the approximate deviation interval

$$1 - \frac{4r^3}{n^3(n-1)^3} \pm \frac{12r^3}{n^3(n-1)^3} \sqrt{\frac{N-n}{Nn}}.$$

## References

- Abelson, R. P. and M. J. Rosenberg  
 1958 Symbolic psycho-logic: A model of attitudinal cognition, *Behavioral Science* 3:1 - 13.
- Cartwright, D. and F. Harary  
 1956 Structural balance: A generalization of Heider's theory, *Psychological Revue* 63:277 - 293.  
 1968 On colorings of signed graphs, *Elemente der Mathematik* 23:85 - 89.  
 1970 Ambivalence and indifference in generalizations of structural balance, *Behavioral Science* 15:497 - 513.
- Frank, O.  
 1971 *Statistical Inference in Graphs*, Research Institute of National Defence, Stockholm.  
 1977a Estimation of graph totals, *Scandinavian Journal of Statistics* 4:81 - 89.  
 1977b A note on Bernoulli sampling in graphs and Horvitz-Thompson estimation, *Scandinavian Journal of Statistics* 4:178 - 180.  
 1977c Survey sampling in graphs, *Journal of Statistical Planning and Inference* 1:235 - 264.  
 1978a Estimation of the number of connected components in a graph by using a sampled sub-graph, *Scandinavian Journal of Statistics* 5:177 - 188.  
 1978b Sampling and estimation in large social networks. *Social Networks* 1:91 - 101.  
 1978c Inferences concerning cluster structure, *Compstat 1978, Proceedings in Computational Statistics, 3rd Symposium*, Leiden 1978, edited by L.C.A. Corsten and J. Hermans, Physica-Verlag, Wien.  
 1979a Estimation of population totals by use of snowball samples, *Perspectives on Social Network Research*, edited by P. Holland and S. Leinhardt, Academic Press, New York.  
 1979b Moment properties of subgraph counts in stochastic graphs, *Annals of the New York Academy of Sciences*, 319: 207 - 218.  
 1979c Estimating a graph from triad counts, *Journal of Statistical Computation and Simulation*, 9: 31 - 46.  
 1979d Transitivity in stochastic graphs and digraphs, *Journal of Mathematical Sociology* (submitted).
- Harary, F.  
 1953 On the notion of balance of a signed graph, *Michigan Mathematical Journal* 2:143 - 146.  
 1955 On local balance and N-balance in signed graphs, *Michigan Mathematical Journal* 3:37 - 41.  
 1957 Structural duality, *Behavioral Science* 2:255 - 265.



- 1959 On the measurement of structural balance, *Behavioral Science* 4:316 - 323.
- 1960 A matrix criterion for structural balance, *Naval Research Logistics Quarterly* 7:195 - 199.
- Heider, F.
- 1946 Attitudes and cognitive organization. *Journal of Psychology* 21:107 - 112.
- 1958 *The Psychology of Interpersonal Relations*, Wiley, New York.
- Holland, P. and S. Leinhardt
- 1975 Local structure in social networks, *Social Methodology 1976*, edited by D. Heise, Jossey-Bass, San Francisco.
- 1979 Structural sociometry, *Perspectives on Social Network Research*, edited by P. Holland and S. Leinhardt, Academic Press, New York.
- Zajonc, R.
- 1968 Cognitive theories in social psychology, *Handbook of Social Psychology*, edited by G. Lindzey and E. Aronson, 319 - 411.