Balance in Random Signed Graphs

A. EL Maftouhi, Y. Manoussakis
Laboratoire de Recherche en Informatique
Université Paris Sud 11, Bât. 650
91405 Orsay Cedex, France
hakim.maftouhi@orange.fr, yannis@lri.fr
and
O. Megalakaki
Faculté de Philosophie, Université de Picardie
Chemin du Thil, 8025 Amiens Cedex 01 France
olga.megalakaki@u-picardie.fr

Abstract

By extending Heider's and Cartwright-Harary's theory of balance in deterministic social structures, we study the problem of balance in social structures where relations between individuals are random. An appropriate model for representing such structures is the so called random signed graphs $G_{n,p,q}$ defined as follows. Given a set of n vertices and fixed numbers p and q, $0 < p+q < 1$, then, between each pair of vertices, there exists either a positive edge with probability p, or a negative edge with probability q, or there is no edge with probability $1-p-q$.

We first show that, almost always (*i.e.* with probability tending to 1 as $n \longrightarrow \infty$), the random signed graph $G_{n,p,q}$ is unbalanced. Subsequently we estimate the maximum order of a balanced induced subgraph in $G_{n,p,p}$, and show that its order achieves only a finite number of values. Next, we study the asymptotic behavior of the degree of balance and give upper and lower bounds for the line-index of balance. Finally, we study the threshold function of balance, e.g., a function $p_0(n)$ such that if $p \gg p_0(n)$, then almost always the random signed graph $G_{n,p,p}$ is unbalanced, else it is almost always balanced.

1 Introduction and terminology

Within the rapid growth of the Internet and the Web, and in the ease with which global communication now takes place, connectedness took an important place in modern society. Global phenomena, involving social networks, incencitives and the behavior of people based on the links that connect us appear in a regular manner. Motivated by these developements, there is a growing multidisciplinary interest to understand how highly connected systems operate [3]. In our discussion here, we consider social networks settings with both positive and negative effects. Some realtions are friendly, but others are antagonistic or hostile. In such a context, let P define a *population* of n individuals. Given a symmetric relationship between individuals in P, the simplest approach to study the behavior of such a population is to consider a graph G in which the vertices represent the individuals, and there exists an edge between two vertices x and y in G if and only if the corresponding individuals are in relation in P. In social sciences we often deal with relations of opposite content, e.g., "love"-"hatred", "likes"-"dislikes", "tells truth to"-"lies to" etc. In common use opposite relations are termed positive and negative relations. A signed graph is one in which relations between entities may be of various types in contrast to an unsigned graph where all relations are of the same type. In signed graphs edge-coloring provides an elegant and uniform representation of the various types of relations where every type of relation is represented by a distinct color.

In the case where precisely one relation and its opposite are under consideration, then instead of two colors, the signs $+$ and $-$ are assigned to the edges of the corresponding graph in order to distinguish a relation from its opposite. Formally, a signed graph is a graph $G = (V,\ E)$ together with a function $f :\ E \longrightarrow \ \{+, -\}$, which associates each edge with the sign $+$ or $-$. In such a signed graph, a subset H of $E(G)$ is said to be positive if it contains an even number of negative edges, otherwise is said to be negative. A signed graph G is *balanced* if each cycle of G is positive. Otherwise it is unbalanced.

The theory of balance goes back to Heider (1946) [10] who asserted that a social system is balanced if there is no tension and that unbalanced social structures exhibit a tension resulting in a tendency to change in the direction of balance.

Since this first work of Heider, the notion of balance has been extensively studied by many mathematicians and psychologists. For a survey see [15].

In 1956, Cartwright and Harary [2] provided a mathematical model for balance through graphs. Their cornerstone result states that a signed graph is balanced if and only if in each cycle the number of negative edges is even. The following theorem of Harary gives an equivalent definition of a balanced signed graph.

Theorem 1.1 (Harary [7]). *A signed graph is balanced if and only if its vertex set can be partitioned into two classes (one of the two classes may be empty) so that every edge joining vertices within a class is positive and every edge joining vertices between classes is negative.*

In 1958, Morissette [12] introduced the notion of "*degree of balance*", a measure of relative balance by which one can decide whether one unbalanced structure is more balanced than another one. Cartwright and Harary [2] suggested an approximation of the degree of balance by studying the rather naive ratio $\rho = X^+/X$ of the number X^+ of positive cycles to the total number X of cycles. Clearly, ρ lies between 0 and 1. Later, Flament [5], Cartwright and Harary [2], Taylor [16] and Norman-Roberts [13] observed that cycles of different lengths contribute differently to balance, with longer cycles being less important than shorter ones. Thus, one natural way to speak of *relative m-balance* is to use the ratio of the number of positive cycles of length at most m to the total number of cycles of length at most m. Norman and Roberts [13] proposed to study relative balance by using the ratio

$$\frac{\sum\limits_{m\geq 3} f(m)X_m^+}{\sum\limits_{m\geq 3} f(m)(X_m^+ + X_m^-)},$$

where X_m^+ (X_m^-) denotes the number of positive (negative) cycles of length m and $f(m)$ is a monotone decreasing function which weights the relative importance of cycles of length m.

In another rather different approach, balance is measured by counting the smallest number δ of edges whose inversion of signs would result in a balanced signed graph. The parameter δ

is called the *line-index of balance*. An interesting result concerning the line-index of balance can be found in [8], where the following result has been proved.

Theorem 1.2 (Harary [8]). *The line-index δ of balance of a signed graph G is the smallest number of edges whose removal from G results in balance.*

For other results on this measure the reader is refereed to [9] and the survey paper of Taylor [16].

In this work we deal with a probabilistic model where we assume that relations between individuals are random (see also [6]). A good mathematical model for representing such random social structures is the so called random signed graph $G_{n,p,q}$ which we introduce here as follows. Let p, q be fixed, $0 < p + q < 1$. Given a set of n vertices, $V = \{1, \cdots, n\}$, between each pair of distinct vertices x and y there is either a positive edge with probability p or a negative edge with probability q, or else there is no edge at all with probability $1-(p+q)$. The edges between different pairs of vertices are chosen independently. Another way to define the random signed graph $G_{n,p,q}$ is as follows. Define first the random (non-signed) graph $\tilde{G}_{n,p,q}$ ($\tilde{G}_{n,p,q}$ has the same probability distribution as the standard random graph $G_{n,p+q}$ with edge probability $p+q$). Next, for any fixed pair $\{x, y\}$ of vertices of V, assign

$$\Pr\left[\ \{x,y\}\ \textit{is positive in}\ G_{n,p,q}\ \middle|\ \{x,y\} \in E(\tilde{G}_{n,p,q})\right] = \frac{p}{p+q}$$

and

$$\Pr\left[\ \{x,y\}\ \textit{is negative in}\ G_{n,p,q}\ \middle|\ \{x,y\} \in E(\tilde{G}_{n,p,q})\right] = \frac{q}{p+q}.$$

In other words, $G_{n,p,q}$ can be considered as the random variable on the set of the signed graphs on n vertices whose probability distribution is given by

$$\Pr\left[G_{n,p,q} = G_0\right] = p^m q^k (1-p-q)^{\binom{n}{2}-m-k}.$$

where G_0 is a fixed signed graph with m positive edges and k negative edges.

Throughout this paper, if $\mathcal{P}$ is a graph property then the expression "*almost always*" $G_{n,p,q}$ satisfies $\mathcal{P}$, means " *with probability tending to* 1 *as* $n \longrightarrow \infty$", $G_{n,p,q}$ satisfies $\mathcal{P}$.

In this work we study the aforementioned measures of balance in the case of random signed graphs. In particular, in the next section we show that, almost always, the random signed graph $G_{n,p,q}$ is unbalanced. Then we estimate the maximum order $\beta = \beta(G_{n,p,p})$ of a balanced induced subgraph in $G_{n,p,p}$, and show that, almost always, β achieves only a finite number of values.

In Section 3, we study relative m-balance in $G_{n,p,p}$, and prove that for a fixed integer m, the ratio $\dfrac{X_m^+}{X_m^+ + X_m^-}$ tends to $\dfrac{1}{2}$ with probability tending to 1 as $n \longrightarrow \infty$.

In Section 4, we derive estimates of the upper and lower bounds for the line-index of balance.

Finally, in Section 5 we study the threshold function of balance, that is a function $p_0(n)$ such that if $p \gg p_0(n)$, then almost no signed graph is balanced, and if $p \ll p_0(n)$, then almost every signed graph is balanced.

Throughout this paper we shall use the following notations and definitions. Let $G = G(V, E)$ be a signed graph with vertex set V and edge set $E = E(G)$. We shall denote by $\tilde{G}$ the underlying simple graph obtained from G by ignoring the signs of its edges.

Let $\mathfrak{C}_m = \mathfrak{C}_m(K_n)$ denote the set of all possible cycles of length m in the complete graph K_n on n vertices. Clearly $|\mathfrak{C}_m| = \frac{(m-1)!}{2}\binom{n}{m}$.

If C_m is an element of $\mathfrak{C}_m$, then the notation $\tilde{G}_{n,p,p} \supseteq C_m$ means that the cycle C_m is contained in $\tilde{G}_{n,p,p}$. We let X_m denote the number of cycles of length m contained in the random graph $\tilde{G}_{n,p,p}$

$$X_m = \sum_{C_m \in \mathfrak{C}_m} 1_{\{\tilde{G}_{n,p,p} \supseteq C_m\}}.$$

X_m is also the total number of (positive and negative) cycles of length m in the random signed graph $G_{n,p,p}$. Furthermore, X_m^+ (X_m^-) denotes the number of positive (negative) cycles of length m in $G_{n,p,p}$.

We observe that in our probabilistic model the random signed graph $G_{n,p,q}$ is almost always connected and it contains at least one cycle of an arbitrary length (see [14, page 14]).

2 The maximum order of a balanced induced subgraph

In view of Theorem 2.2 we prove the following lemma.

Lemma 2.1. *Let H be a fixed set of h distinct pairs of vertices of $G_{n,p,q}$.*

$$\Pr\left[H \textit{ is positive in } G_{n,p,q} \,\middle|\, H \subseteq E(\tilde{G}_{n,p,q})\right] = \frac{1}{2}\left[1 + \left(\frac{p-q}{p+q}\right)^h\right]$$

and

$$\Pr\left[H \textit{ is negative in } G_{n,p,q} \,\middle|\, H \subseteq E(\tilde{G}_{n,p,q})\right] = \frac{1}{2}\left[1 - \left(\frac{p-q}{p+q}\right)^h\right].$$

Proof. Let H be a fixed set of h pairs of vertices. Then

$$\begin{aligned} p_1 &= \Pr\left[H \textit{ is positive in } G_{n,p,q} \,\middle|\, H \subseteq E(\tilde{G}_{n,p,q})\right] \\ &= \sum_{i \textit{ even}} \Pr\left[|H^-| = i\right], \end{aligned}$$

where $|H^-|$ is the number of negative edges in H. Thus

$$p_1 = \frac{1}{(p+q)^h} \sum_{i \textit{ even}} \binom{h}{i} q^i p^{h-i}.$$

Similarly,

$$p_2 = \Pr[H \textit{ is negative in } G_{n,p,q} \,|\, H \subseteq E(\tilde{G}_{n,p,q})] \quad = \quad \frac{1}{(p+q)^h} \sum_{i \textit{ odd}} \binom{h}{i} q^i p^{h-i}.$$

We obtain the following system of equations

$$\begin{cases} p_1 + p_2 &= 1 \\ p_1 - p_2 &= \left[\dfrac{p-q}{p+q}\right]^h. \end{cases}$$

By solving this system, we obtain the desired expressions for p_1 and p_2. □

Theorem 2.2. *Let p and q be fixed positive real numbers, $0 < p+q < 1$. Then, almost always, $G_{n,p,q}$ is unbalanced.*

Proof. Let $\mathfrak{T}$ denote a maximum set of disjoint-edge triangles in the complete graph K_n. To prove the theorem it suffices to show that, almost always, $G_{n,p,q}$ contains a negative triangle from $\mathfrak{T}$.

Clearly $|\mathfrak{T}| \geq \left\lfloor \frac{n}{3} \right\rfloor$. Let T be a fixed element of $\mathfrak{T}$. We have

$$\Pr\left[T \subseteq \tilde{G}_{n,p,q} \ \textit{and} \ T \ \textit{is} \ \textit{negative}\right] = \Pr\left[T \ \textit{is} \ \textit{negative} \ \middle| \ T \subseteq \tilde{G}_{n,p,q}\right] \times \Pr\left[\ T \subseteq \tilde{G}_{n,p,q}\right]$$

Using Lemma 2.1, we get

$$\begin{aligned}
\Pr\left[T \subseteq E(\tilde{G}_{n,p,q}) \ \textit{and} \ T \ \textit{is} \ \textit{negative}\right] &= \frac{1}{2}\left[1 - \left(\frac{p-q}{p+q}\right)^3\right](p+q)^3 \\
&= \frac{1}{2}\left[(p+q)^3 - (p-q)^3\right]
\end{aligned}$$

Thus, the probability that $G_{n,p,q}$ contains a negative triangle from $\mathfrak{T}$ is at least

$$1 - \left(1 - \frac{1}{2}\left[(p+q)^3 - (p-q)^3\right]\right)^{\lfloor \frac{n}{3} \rfloor}.$$

As p and q are fixed, this last expression tends to 1 as $n \longrightarrow \infty$.

$\square$

A natural problem which arises from this theorem is to derive estimates of the maximum order, denoted by $\beta = \beta(G_{n,p,p})$, of a balanced induced subgraph in $G_{n,p,p}$ where p is fixed. The following theorem shows that, almost always, β achieves only a finite number of values. More precisely, let $d(n)$ be the function defined by

$$d(n) = 2\log_{\frac{1}{1-p}}(n) - 2\log_{\frac{1}{1-p}}\log_{\frac{1}{1-p}}(n) + 1 + 2\log_{\frac{1}{1-p}}(\frac{e}{2}).$$

Theorem 2.3. *Let $\epsilon > 0$ be fixed. Let p be fixed, $0 < 2p < 1$. Then*

$$\Pr\left[\lfloor d(n) - \epsilon \rfloor \leq \beta(G_{n,p,p}) \leq \lfloor d(n) + 2\log_{\frac{1}{1-p}} 2 + \epsilon \rfloor\right] \longrightarrow 1 \quad \textit{as} \quad n \longrightarrow \infty.$$

Proof. Since each induced subgraph of $G_{n,p,p}$ without negative edges is balanced, we obviously have

$$\beta(G_{n,p,p}) \geq \alpha(G_{n,p}),$$

where $\alpha(G_{n,p})$ denotes the independence number of the random graph $G_{n,p}$. Using the following result due to Matula[11] (see also [1, page 251]

$$\Pr\left[\lfloor d(n) - \epsilon \rfloor \leq \alpha(G_{n,p}) \leq \lfloor d(n) + \epsilon \rfloor\right] \longrightarrow 1 \quad \textit{as} \quad n \longrightarrow \infty.$$

we get the lower bound for $\beta(G_{n,p,p})$

$$\Pr\left[\lfloor d(n) - \epsilon \rfloor \leq \beta(G_{n,p,p})\right] \longrightarrow 1 \quad \text{as} \quad n \longrightarrow \infty.$$

To conclude the proof, it remains to show that there exists no balanced induced subgraph of order $> \lfloor d(n) + 2\log_{\frac{1}{1-p}} 2 + \epsilon \rfloor$; that is

$$\Pr\left[\beta(G_{n,p,p}) > \lfloor d(n) + 2\log_{\frac{1}{1-p}} 2 + \epsilon \rfloor\right] \longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty.$$

Let N_r be the number of sets of r vertices whose induced subgraph is balanced. By using Markov inequality

$$\Pr\left[N_r \geq 1\right] \leq E(N_r),$$

it suffices to prove that, for $r > d(n) + 2\log_{\frac{1}{1-p}} 2 + \epsilon$, $E(N_r) \longrightarrow 0$ as $n \longrightarrow \infty$, which implies that $\Pr[N_r = 0] \longrightarrow 1$ as $n \longrightarrow \infty$.

Let S be a fixed set of r vertices of $G_{n,p,p}$. By Theorem 1.1, the subgraph induced by S is balanced if and only if S can be partitioned into two classes V_1 and V_2 such that

(i) the subgraph induced by V_1 (resp. V_2) contains no negative edge,

(ii) there is no positive edge between V_1 and V_2 .

The probability that a given bipartition $\{V_1, V_2\}$ of S satisfies simultaneously conditions (i) and (ii) is

$$(1-p)^{\frac{r(r-1)}{2}}.$$

The probability that there exists a partition of S satisfying the above conditions is smaller than

$$2^{r-1}(1-p)^{\frac{r(r-1)}{2}}.$$

Thus

$$E(N_r) \leq 2^{r-1}\binom{n}{r}(1-p)^{\frac{r(r-1)}{2}}.$$

Using Stirling's formula, we get

$$E(N_r) \leq \frac{1}{2\sqrt{2\pi r}}\left[\frac{2en(1-p)^{\frac{(r-1)}{2}}}{r}\right]^r.$$

Hence, $E(N_r) \longrightarrow 0$ if, for large n, we have

$$\frac{2en(1-p)^{\frac{(r-1)}{2}}}{r} \leq 1. \tag{1}$$

Set

$$f(r) = \frac{2en(1-p)^{\frac{(r-1)}{2}}}{r}.$$

Let ϵ be a fixed real number $\epsilon > 0$. Since f is a monotone decreasing function, inequality (1) will certainly be true, for $r > d(n) + 2\log_{\frac{1}{1-p}} 2 + \epsilon$, if

$$f\left(d(n) + 2\log_{\frac{1}{1-p}} 2 + \epsilon\right) \leq 1. \tag{2}$$

A straightforward computation shows that (2) is equivalent to

$$\frac{2(1-p)^{\frac{\epsilon}{2}}\log_{\frac{1}{1-p}}(n)}{d(n)+2\log_{\frac{1}{1-p}}2+\epsilon}\leq 1. \tag{3}$$

On replacing $d(n)+2\log_{\frac{1}{1-p}}2+\epsilon$ by its lower bound $2\log_{\frac{1}{1-p}}(n)-2\log_{\frac{1}{1-p}}\log_{\frac{1}{1-p}}(n)$, we see that (3) is satisfied if

$$\frac{2(1-p)^{\frac{\epsilon}{2}}\log_{\frac{1}{1-p}}(n)}{2\log_{\frac{1}{1-p}}(n)-2\log_{\frac{1}{1-p}}\log_{\frac{1}{1-p}}(n)}\leq 1.$$

It is not hard to see that the above condition is asymptotically true, since $(1-p)^{\frac{\epsilon}{2}}<1$. This completes the proof. □

3 The degree of balance

In Theorem 3.3 formulated later in this section, we study relative m-balance by using the ratio $\delta = X_m^+/X_m$ of the number X_m^+ of positive cycles of length m to the total number X_m of cycles of length m in $G_{n,p,p}$. In view of the proof of Theorem 3.3, we first prove two lemmas which could be interesting on their own.

Lemma 3.1. *Let $H_1, H_2, \cdots, H_t$ be t fixed distinct sets of pairs of vertices, $t \geq 2$. Let $H = \bigcup_{i=1}^{t} H_i$. Then, the events*

$$\left\{ H_i \text{ } is \text{ positive } in \text{ } G_{n,p,p} \,\middle|\, H \subseteq E(\tilde{G}_{n,p,p}) \right\}, \quad for \;\; i = 1, \cdots, t\ ,$$

are pairwise independent.

Proof. We denote by P_H the conditional probability given $\{H \subseteq E(\tilde{G}_{n,p,q}\}$. Let H_i and H_j be two distinct elements of $\{H_1, \cdots, H_t\}$. We have to prove that

$$P_H\Big[H_i \textit{ and } H_j \textit{ are both positive}\Big] = P_H\Big[H_i \textit{ is positive}\Big] \times P_H\Big[H_j \textit{ is positive}\Big].$$

Since, by Lemma 2.1, each probability in the right side of the above expression is equal to $\frac{1}{2}$, it suffices to show that

$$P_H\Big[H_i \textit{ and } H_j \textit{ are both positive}\Big] = \frac{1}{4}.$$

Observe first that the statement is trivially true when H_i and H_j are disjoint sets.

Suppose now that one of the two sets is contained in the other, for example $H_i \subset H_j$. Then

$$P_H\Big[H_i \textit{ and } H_j \textit{ are both positive}\Big] = \\ P_H\Big[H_i \textit{ is positive}\Big] \times P_H\Big[H_j \setminus H_i \textit{ is positive}\Big] = \frac{1}{4}$$

Consider now the case where $H_i \cap H_j \neq \emptyset$, and none of the two sets is contained in the other. Then, clearly H_i and H_j are both positive if and only if either each of $H_i \setminus H_j,\ H_i \cap H_j,\ H_j \setminus H_i$ is positive or each of $H_i \setminus H_j,\ H_i \cap H_j,\ H_j \setminus H_i$ is negative. Thus

$$P_H[H_i \textit{ and } H_j \textit{ are both positive }]$$

$$= P_H\Big[H_i \setminus H_j \textit{ is positive}\Big] \times P_H\Big[H_j \setminus H_i \textit{ is positive}\Big] \times P_H\Big[H_i \cap H_j \textit{ is positive}\Big]$$

$$+ P_H\Big[H_i \setminus H_j \textit{ is negative}\Big] \times P_H\Big[H_j \setminus H_i \textit{ is negative}\Big] \times P_H\Big[H_i \cap H_j \textit{ is negative}\Big].$$

By Lemma 2.1, each probability in the right side of the above equality is equal to $1/2$. Thus

$$P_H\Big[H_1 \textit{ and } H_2 \textit{ are both positive}\Big] = \frac{1}{4}.$$

$\square$

Lemma 3.2. *Let m be a fixed integer, $0 \leq m \leq n$. Let X_m denote the total number of cycles of length m in $G_{n,p,p}$. Then, for any arbitrarily small $\epsilon > 0$,*

$$\Pr\Big[\Big|X_m - E(X_m)\Big| \geq \epsilon E(X_m)\Big] \leq \frac{4m^{2m+3}}{\epsilon^2 n (2p)^m}.$$

Proof.

Expectation of X_m

Clearly

$$E(X_m) = \frac{(m-1)!}{2}\binom{n}{m}(2p)^m.$$

Since m is fixed, we have

$$\frac{(m-1)!}{2}\binom{n}{m} \sim \frac{n^m}{2m}.$$

Thus

$$E(X_m) \sim \frac{n^m}{2m}(2p)^m. \tag{4}$$

Variance of X_m

$$\begin{aligned}
E(X_m^2) &= E\left[\sum_{C_m \in \mathfrak{C}_m} 1_{\{G_{n,p,p} \supseteq C_m\ \}}\right]^2 \\
&= \sum_{C_m, C'_m \in \mathfrak{C}_m} \Pr\Big[G_{n,p,p} \supseteq C_m \text{ and } G_{n,p,p} \supseteq C'_m\Big] \\
&= \sum_{k=0}^{m}\left[\sum_{|C_m \cap C'_m| = k} \Pr\Big[G_{n,p,p} \supseteq C_m \text{ and } G_{n,p,p} \supseteq C'_m\Big]\right],
\end{aligned} \tag{5}$$

where, for a fixed k, the first sum is considered over all cycles having precisely k vertices in common.

For $k = 0$,

$$\sum_{|C_m \cap C'_m|=0} \Pr\Big[G_{n,p,p} \supseteq C_m \text{ and } G_{n,p,p} \supseteq C'_m\Big]$$
$$= \left[\frac{(m-1)!}{2}\right]^2 \binom{n}{m}\binom{n-m}{m}(2p)^{2m}$$
$$\sim \frac{n^{2m}(2p)^{2m}}{4m^2} \quad . \tag{6}$$

For $k \geq 1$,

$$\sum_{|C_m \cap C'_m|=k} \Pr\Big[G_{n,p,p} \supseteq C_m \text{ and } G_{n,p,p} \supseteq C'_m\Big]$$
$$\leq (m!)^2 \binom{n}{m}\binom{m}{k}\binom{n-m}{m-k}(2p)^{2m-k}$$
$$\leq m^{2k} n^{2m-k}(2p)^{2m-k} \quad . \tag{7}$$

By (5), (6) and (7),

$$\begin{aligned} E(X_m^2) &\leq \frac{n^{2m}(2p)^{2m}}{4m^2} + \sum_{k=1}^{m} m^{2k} n^{2m-k}(2p)^{2m-k} \\ &\leq \frac{n^{2m}(2p)^{2m}}{4m^2} + \sum_{k=1}^{m} m^{2m} n^{2m-1}(2p)^{m} \\ &\leq \frac{n^{2m}(2p)^{2m}}{4m^2} + m^{2m+1} n^{2m-1}(2p)^{m}. \end{aligned} \tag{8}$$

From (4) and (8) we obtain

$$\frac{E(X_m^2)}{E^2(X_m)} \leq 1 + \frac{4m^{2m+3}}{n(2p)^m}.$$

Thus

$$\frac{\operatorname{var}(X_m)}{E^2(X_m)} \leq \frac{4m^{2m+3}}{n(2p)^m}.$$

Using Chebyshev's inequality

$$\Pr\Big[\Big|X_m - E(X_m)\Big| \geq \epsilon E(X_m)\Big] \leq \frac{\operatorname{var}(X_m)}{\epsilon^2 E^2(X_m)},$$

we conclude the proof.

The following theorem concerns the relative m-balance in $G_{n,p,p}$. In order to formulate it, let us define the random variable $\rho(m)$ as follows

$$\rho(m) = \begin{cases} \dfrac{X_m^+}{X_m} & if \quad X_m \neq 0 \\ \\ \dfrac{1}{2} & if \quad X_m = 0 \end{cases}$$

Theorem 3.3. *Let m be a fixed integer, $0 \leq m \leq n$. Then, almost always in $G_{n,p,p}$, we have $\rho(m) \longrightarrow \dfrac{1}{2}$.*

Proof. Let $\mathfrak{C}_m$ be, as defined in Section 1, the set of cycles each of length m in the complete graph K_n on n vertices. Let k be a fixed integer, $0 \leq k \leq \dfrac{(m-1)!}{2}\dbinom{n}{m}$. Let $\mathfrak{C}_{m,k}$ be a fixed subset of $\mathfrak{C}_m$ of cardinality k. We denote by $X_{m,k}^+$ the random variable X_m^+ conditioned by the event $\{\mathfrak{C}_m(\tilde{G}_{n,p,p}) = \mathfrak{C}_{m,k}\}$,

$$X_{m,k}^+ = \left\{ X_m^+ \,\middle|\, \mathfrak{C}_m(\tilde{G}_{n,p,p}) = \mathfrak{C}_{m,k} \right\} ,$$

where $\mathfrak{C}_m(\tilde{G}_{n,p,p})$ denotes the set of cycles of length m contained in $\tilde{G}_{n,p,p}$. $X_{m,k}^+$ can be expressed as follows

$$X_{m,k}^+ = \sum_{C_m \in \mathfrak{C}_{m,k}} 1_{\{C_m \ is \ positive \,|\, \mathfrak{C}_m(\tilde{G}_{n,p,p}) = \mathfrak{C}_{m,k}\}}.$$

By Lemma 2.1, for each $C_m \in \mathfrak{C}_{m,k}$, we have

$$\Pr\left[C_m \ is \ positive \,\middle|\, \mathfrak{C}_m(\tilde{G}_{n,p,p}) = \mathfrak{C}_{m,k} \right] = \frac{1}{2}.$$

Thus

$$E(X_{m,k}) = \frac{k}{2}. \tag{9}$$

Now, by Lemma 3.1, for $C_m,\ C'_m \in \mathfrak{C}_m$, $C_m \neq C'_m$, the events $\left\{C_m \ is \ positive \,\middle|\, \mathfrak{C}_m(\tilde{G}_{n,p,p}) = \mathfrak{C}_{m,k}\right\}$ and $\left\{C'_m \ is \ positive \,\middle|\, \mathfrak{C}_m(\tilde{G}_{n,p,p}) = \mathfrak{C}_{m,k}\right\}$ are independent. Thus

$$\mathrm{var}(X_{m,k}^+) = \frac{k}{4}. \tag{10}$$

Expectation of $\rho(m)$

Since $\{ \mathfrak{C}_m(\tilde{G}_{n,p,p}) = \mathfrak{C}_{m,k}\} \subseteq \{X_m = k\}$, it follows that

$$E(X_m^+ \,\Big|\, X_m = k) = \sum_{\mathfrak{C}_{m,k}} E(X_{m,k}^+) \times \Pr\left[\mathfrak{C}_m(\tilde{G}_{n,p,p}) = \mathfrak{C}_{m,k} \,\middle|\, X_m = k\right] ,$$

where the sum is over all the subsets $\mathfrak{C}_{m,k}$ of $\mathfrak{C}_m$. Using (9), we obtain

$$E(X_m^+ \,\Big|\, X_m = k) = \frac{k}{2} \sum_{\mathfrak{C}_{m,k}} \Pr\left[\mathfrak{C}_m(\tilde{G}_{n,p,p}) = \mathfrak{C}_{m,k} \,\middle|\, X_m = k\right] ,$$

Since the above sum is equal to 1, we get

$$E(X_m^+ \;\Big|\; X_m = k) = \frac{k}{2}.$$

It follows that, for $k \geq 1$, we have

$$E(\rho(m) \;\Big|\; X_m = k) = \frac{1}{2}.$$

From the definition of $\rho(m)$, we have, for $k = 0$

$$E(\rho(m) \;\Big|\; X_m = 0) = \frac{1}{2}.$$

Thus

$$E(\rho(m)) = \frac{1}{2}.$$

Variance of $\rho(m)$

Using again the fact that $\{\ \mathfrak{C}_m(\tilde{G}_{n,p,p}) = \mathfrak{C}_{m,k}\} \subseteq \{X_m = k\}$, we get

$$var(X_m^+ \;\Big|\; X_m = k) = \sum_{\mathfrak{C}_{m,k}} var(X_{m,k}^+) \times \Pr\left[\mathfrak{C}_m(\tilde{G}_{n,p,p}) = \mathfrak{C}_{m,k} \;\middle|\; X_m = k\right].$$

Equality (10) gives

$$\mathrm{var}\left[X_m^+ \;\Big|\; X_m = k\right] = \frac{k}{4}.$$

Hence, for $1 \leq k \leq \dfrac{(m-1)!}{2}\dbinom{n}{m}$,

$$\mathrm{var}\left[\rho(m) \;\Big|\; X_m = k\right] = \frac{1}{4k},$$

and, from the definition of $\rho(m)$, we have for $k = 0$

$$\mathrm{var}\left[\rho(m) \;\middle|\; X_m = 0\right] = 0.$$

Thus,

$$\begin{aligned}
\mathrm{var}[\rho(m)] &= \sum_{k=0}^{\frac{(m-1)!}{2}\binom{n}{m}} \mathrm{var}\left[\rho(m) \;\middle|\; X_m = k\right] \times \Pr\left[X_m = k\right] \\
&= \sum_{k=1}^{\frac{(m-1)!}{2}\binom{n}{m}} \frac{1}{4k} \Pr\left[X_m = k\right]. \qquad (11)
\end{aligned}$$

Let ϵ be arbitrarily small positive real number. Then

$$\Pr\Big[X_m = k\Big] =$$

$$\Pr\Big[X_m = k \,\Big|\, |X_m - E(X_m)| > \epsilon E(X_m)\Big] \times \Pr\Big[|X_m - E(X_m)| > \epsilon E(X_m)\Big]$$

$$+ \Pr\Big[X_m = k \,\Big|\, |X_m - E(X_m)| \le \epsilon E(X_m)\Big] \times \Pr\Big[|X_m - E(X_m)| \le \epsilon E(X_m)\Big].$$

Since $\frac{(m-1)!}{2}\binom{n}{m} \le n^m$, and by (11), we have

$$\operatorname{var}\Big[\rho(m)\Big] \le \left[\sum_{k=0}^{n^m} \frac{1}{4k}\right] \times \Pr\Big[|X_m - E(X_m)| > \epsilon E(X_m)\Big] + \sum_{\frac{(1-\epsilon)n^m(2p)^m}{2m} \le k \le \frac{(1+\epsilon)n^m(2p)^m}{2m}} \frac{1}{4k}.$$

Lemma 3.2 gives

$$\operatorname{var}[\rho(m)] \le \left[\sum_{k=0}^{n^m} \frac{1}{k}\right] \frac{m^{2m+3}}{\epsilon^2 n(2p)^m} + \sum_{\frac{(1-\epsilon)n^m(2p)^m}{2m} \le k \le \frac{(1+\epsilon)n^m(2p)^m}{2m}} \frac{2m}{(1-\epsilon)n^m(2p)^m}.$$

Since $\left[\sum_{k=0}^{n^m} \frac{1}{k}\right] = O(m \log n)$,

$$\operatorname{var}\Big[\rho(m)\Big] \le \frac{m^{2m+4}}{\epsilon^2 n(2p)^m} O(\log n) + \frac{2\epsilon}{1-\epsilon}$$

As ϵ is an arbitrary small positive number, by setting $\epsilon = \frac{1}{\log n}$ in this last inequality, we obtain

$$\operatorname{var}\Big[\rho(m)\Big] = o(1).$$

The end of the proof follows from Chebyshev's inequality. □

4 The line-index of balance

Let us recall that, by Theorem 1.2, the line-index δ of a signed graph is the smallest number of edges whose removal results in balance. In the next theorem we give estimates for the upper and lower bounds of $\delta(G_{n,p,p})$.

Theorem 4.1. *Let ϵ be an arbitrarily small positive number. Then the line-index of balance δ of $G_{n,p,p}$ satisfies*

$$\Pr\left[(1-\epsilon)\frac{n^2p}{2} \le \delta \le (1+\epsilon)\frac{n^2p}{2}\right] \longrightarrow 1$$

as $n \longrightarrow \infty$.

Proof. Let $\{S,T\}$ be a fixed partition of the vertex set of $G_{n,p,p}$. Set $|S| = s$ and $|T| = t$, where $s+t = n$. Let $Y_{S,T}$ be the random variable equal to the number $|E^+(S,T)|$ of positive edges between S and T plus the number $|E^-(S)|$of negative edges in the subgraph induced by S and the number $|E^-(T)|$ of negative edges in the subgraph induced by T,

$$Y_{S,T} = |E^+(S,T)| + |E^-(S)| + |E^-(T)|.$$

It can be easily verified that $Y_{S,T}$ has a binomial distribution with parameters $\frac{n(n-1)}{2}$ and p. Thus

$$E(Y_{S,T}) = \frac{n(n-1)p}{2} \sim \frac{n^2p}{2}.$$

For any $\epsilon > 0$, Chernoff's bounds give

$$\Pr\left[\left|Y_{S,T} - \frac{n^2p}{2}\right| > \frac{\epsilon n^2p}{2}\right] \le \exp\left[-\frac{\epsilon^2 n^2 p}{6}\right]. \tag{12}$$

In [4] it has been proved that

$$\delta = \min_{\{S,T\}} Y_{S,T},$$

where the minimum is over all the partitions $\{S,T\}$ of the vertex set of $G_{n,p,p}$.
Since

$$\Pr\left[\ for\ all\ \{S,T\}\ ,\ (1-\epsilon)\frac{n^2p}{2} \le Y_{S,T} \le (1+\epsilon)\frac{n^2p}{2}\right] \le \Pr\left[\ (1-\epsilon)\frac{n^2p}{2} \le \delta \le (1+\epsilon)\frac{n^2p}{2}\right],$$

it follows that

$$\begin{aligned}
&\Pr\left[(1-\epsilon)\frac{n^2p}{2} \le \delta \le (1+\epsilon)\frac{n^2p}{2}\right] \\
&\qquad \ge 1 - \Pr\left[\ for\ some\ \{S,T\}\ ,\ \left|Y_{S,T} - \frac{n^2p}{2}\right| > \frac{\epsilon n^2p}{2}\right] \\
&\qquad\qquad \ge 1 - 2^n \Pr\left[\left|Y_{S,T} - \frac{n^2p}{2}\right| > \frac{\epsilon n^2p}{2}\right].
\end{aligned}$$

Using (12) we obtain

$$\Pr\left[\ (1-\epsilon)\frac{n^2p}{2} \le \delta \le (1+\epsilon)\frac{n^2p}{2}\right] \ge 1 - 2^n \exp\left[-\frac{\epsilon^2 n^2 p}{6}\right].$$

The right side of the above inequality tends to 1 as $n \longrightarrow \infty$. $\square$

5 The threshold function for balance

We suppose that $p = p(n)$ depends on n. Since by Theorem 2.2, when p is fixed, the random signed graph $G_{n,p,p}$ is almost always unbalanced, the purpose of this section is to determine a function $p_0(n)$ such that

(i) if $p \gg p_0(n)$, then, almost always, $G_{n,p,p}$ is unbalanced, while on the other hand,
(ii) if $p \ll p_0(n)$, then, almost always, $G_{n,p,p}$ is balanced.

Such a function $p_0(n)$ is called a *threshold function* for balance.

Theorem 5.1. *If $pn \longrightarrow 0$, then, almost always, $G_{n,p,p}$ is balanced. If $p \geq \frac{c}{n}$, where $c > 2\log 2$ is a constant, then, almost always, $G_{n,p,p}$ is not balanced.*

Proof. Let X denote the total number of cycles in $G_{n,p,p}$. Clearly

$$E(X) = \sum_{k=3}^{n} \binom{n}{k} \frac{(k-1)!}{2} (2p)^k.$$

Therefore,

$$E(X) \leq \sum_{k=3}^{n} \frac{(2pn)^k}{2k},$$

from which it follows that if $pn \longrightarrow 0$, so does $E(X)$. From Markov's inequality we conclude that, almost always, $G_{n,p,p}$ is acyclic, thus almost every signed graph is balanced.

Suppose now that $p > \frac{c}{n}$. With the notations introduced in the proof of Theorem 4.1, let

$$Z = \sum_{\{S,T\}} \mathbb{1}_{\{Y_{S,T}=0\}},$$

where the sum is over all the bipartitions $\{S,T\}$ of the vertex set of $G_{n,p,p}$. Since

$$\Pr[Y_{S,T} = 0] = (1-p)^{\frac{n(n-1)}{2}} \sim (1-p)^{\frac{n^2}{2}},$$

$$E(Z) \sim 2^n (1-p)^{\frac{n^2}{2}} \sim \exp\left[n(\log 2 - \frac{pn}{2})\right].$$

Thus

$$E(Z) \leq \exp\left[n(\log 2 - \frac{c}{2})\right].$$

$c > 2\log 2$ implies that $E(Z) = o(1)$, and by Markov inequality we conclude that $P[Z = 0] \longrightarrow 1$ as $n \longrightarrow \infty$, thus the end of the proof follows from Theorem 1.1. □

References

[1] B. Bollobás, *Random Graphs*, Academic Press, 1985.

[2] D. Cartwright and F. Harary, Structural Balance: A Generalization of Heider's Theory, *Psychological Review* (**63**) 277-293, 1956.

[3] David Easley, Jon Kleinberg, Networks, Crowds, and Markets: Reasoning About a Highly Connected World, *Cambridge University Press*, 2010.

[4] A. El Maftouhi, *PhD thesis*, University of Paris XI (Orsay), 1994.

[5] G. Flament, Théorie des Graphes et Structures Sociales, *Gauthier-Villars*, Paris, 1965.

[6] O. Frank, F. Harary, Balance in Stochastic Signed Graphs, *Social Networks* (**2**) 155-163, 1979/80.

[7] F. Harary, On the Notion of Balance of a Signed Graph, *Michigan Math. J.* (**2**) 143-146, 1954.

[8] F. Harary, On the Measurement of Structural Balance, *Behavioral Sci.* (**4**) 316-323, 1959.

[9] F. Harary, R.Z. Norman and D. Cartwright, *Structural Models: An Introduction to the Theory of Directed Graphs*, Wiley, New York, 1965.

[10] F. Heider, Attitudes and Cognitive Organisation, *Journal of Psychology* (**21**) 107-112, 1946.

[11] D. W. Matula, The Largest Clique Size in a Random Graph, *Technical Report, Dept. of Computer Science*, Southern Methodist University Dallas, 1976.

[12] J. Morissete, On Experimental Study of the Theory of Structural Balance, *Human Relations* (**11**) 239-254, 1958.

[13] R. Norman and F. A. Roberts, Derivation of a Measure of Relative Balance for Social Structures and a Characterization of Extensive Ratio Systems, *Journal of Mathematical Psychology* (**9**) 66-91, 1978.

[14] E. M. Palmer,*Graphical Evolution*, Wiley, New York, 1985.

[15] F. Roberts, *Graph Theory and its Applications to Problems of Society*, CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM Philadelphia, 1978.

[16] H. F. Taylor, Balance in Small Groups, *Van Nostrand Reinhold*, New York, 1970.