Internet Mathematics Vol. 8, No. 4: 364–380

Balance in Random Signed Graphs

A. EL Maftouhi, Y. Manoussakis, and O. Megalakaki

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 in which relations among individuals are random. An appropriate model for representing such structures is that of 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 a positive edge, a negative edge, or no edge with respective probabilities p, q, $1 - p - q$.

We first show that almost always (i.e., with probability tending to 1 as $n \to \infty$), 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 the random signed graph $G_{n,p,p}$ is almost always unbalanced, and otherwise, it is almost always balanced.

1. Introduction and Terminology

Following the rapid growth of the Internet and the World Wide Web, and in light of the ease with which global communication now takes place, connectedness has taken an important place in modern society. Global phenomena involving social networks, incentives, and the behavior of people based on the links that connect us are omnipresent. Motivated by these developments, there has arisen a growing

ISSN: 1542-7951 print

multidisciplinary interest in understanding how highly connected systems operate [Easley and Kleinberg 10]. In our discussion here, we consider social network settings with both positive and negative effects. Some relations 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 studying 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 the social sciences, we often deal with relations of opposing, or antagonistic, content, such as loves–hates, likes–dislikes, and tells the truth–lies to. In common usage, such opposing relations are sometimes given a moral or ethical value and termed positive and negative. A *labeled graph* is one in which relationships between entities may be of various types, in contrast to an unlabeled graph, in which all relations are of the same type. One way of creating a labeled graph is through edge-coloring, which provides an elegant and uniform representation of the various types of relations, in which every type of relation is represented by a distinct color.

In the case that precisely one relation and its opposite are under consideration, then instead of two colors, the signs $+$ and $-$ can be 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 $f : E \to \{+, -\}$ that 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 and 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, who in [Heider 46] 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 [Roberts 78].

In [Cartwright and Harary 56], the authors provided a mathematical model for balance through graphs. Their foundational 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 54] *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) such* *that every edge joining vertices within a class is positive and every edge joining vertices between classes is negative.*

In [Morissete 58], the notion of *degree of balance* was introduced. It is a measure of relative balance by which one can decide whether one unbalanced structure is more balanced than another. In [Cartwright and Harary 56] there was proposed an approximation of the degree of balance obtained 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, it was observed in [Flament 65, Cartwright and Harary 56, Taylor 70, Norman and Roberts 78] that cycles of different lengths contribute differently to balance, with longer cycles being less important than shorter ones. Thus, it becomes natural to speak of *relative m-balance* as the ratio of the number of positive cycles of length at most m to the total number of cycles of length at most m. It was proposed in [Norman and Roberts 78] to study relative balance using the ratio

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

where X_m^+ (X_m^-) denotes the number of positive (negative) cycles of length m, and $f(m)$ is a monotonically decreasing function that 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 [Harary 59], where the following result was proved.

Theorem 1.2. [Harary 59] *The line index of balance δ 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 referred to [Harary et al. 65] and the survey paper [Taylor 70].

In this work, we deal with a probabilistic model in which we assume that relations between individuals are random (see also [Frank and Harary 80]). 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 $V = \{1, \dots, n\}$, then between each pair of distinct vertices x and y there is a positive edge, a negative edge, or there is no edge at all with respective probabilities p, q, $1 - (p + q)$. 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 (unsigned) 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

$$\Pr\big[\{x,y\} \text{ is positive in } G_{n,p,q} \mid \{x,y\} \in E(\tilde{G}_{n,p,q})\big] = \frac{p}{p+q}$$

and

$$\Pr\big[\{x,y\} \text{ is negative in } G_{n,p,q} \mid \{x,y\} \in E(\tilde{G}_{n,p,q})\big] = \frac{q}{p+q}.$$

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

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

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 "$G_{n,p,q}$ satisfies $\mathcal{P}$ almost always" means "$G_{n,p,q}$ satisfies $\mathcal{P}$ with probability tending to 1 as $n \to \infty$."

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 the random signed graph $G_{n,p,q}$ is almost always 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 β almost always 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 $X_m^+/(X_m^+ + X_m^-)$ tends to $\frac{1}{2}$ with probability tending to 1 as $n \to \infty$. 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, which 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 notation 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 $\tilde{G}_{n,p,p}$:

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

Here 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 arbitrary length (see [Palmer 85, p. 14]).

2. The Maximum Order of a Balanced Induced Subgraph

Toward proving Theorem 2.2 below, we prove the following lemma.

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

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

and

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

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

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

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

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

Similarly,

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

We obtain the following system of equations:

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

By solving this system, we obtain the desired expressions for p_1 and p_2. $\square$

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

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

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

$$\begin{aligned} &\Pr\big[T \subseteq \tilde{G}_{n,p,q} \text{ and } T \text{ is negative}\big] \\ &\quad = \Pr\big[T \text{ is negative} \mid T \subseteq \tilde{G}_{n,p,q}\big] \times \Pr\big[T \subseteq \tilde{G}_{n,p,q}\big]. \end{aligned}$$

Using Lemma 2.1, we get

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

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

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

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

A natural problem that 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 β almost always achieves only a finite number of values. More precisely, let $d(n)$ be the function defined by

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

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

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

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 result

$$\Pr\big[\lfloor d(n)-\epsilon\rfloor \le \alpha(G_{n,p}) \le \lfloor d(n)+\epsilon\rfloor\big] \to 1 \text{ as } n\to\infty,$$

from [Matula 76] (see also [Bollobás 85, p. 251]), we get a lower bound for $\beta(G_{n,p,p})$:

$$\Pr\big[\lfloor d(n)-\epsilon\rfloor \le \beta(G_{n,p,p})\big] \to 1 \text{ as } n\to\infty.$$

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

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

Let N_r be the number of sets of r vertices whose induced subgraph is balanced. Using Markov's inequality

$$\Pr\big[N_r \ge 1\big] \le E(N_r),$$

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

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 the following hold:

(i) The subgraph induced by V_1 and the subgraph induced by V_2 contain 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

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

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

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

Thus

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

Using Stirling's formula, we get

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

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

$$\frac{2en(1-p)^{(r-1)/2}}{r} \leq 1. \tag{2.1}$$

Set

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

Let ϵ be a fixed positive real number. Since f is a monotonically decreasing function, inequality (2.1) will certainly be true for $r > d(n) + 2\log_{\frac{1}{1-p}} 2 + \epsilon$ if

$$f(d(n) + 2\log_{\frac{1}{1-p}} 2 + \epsilon) \leq 1. \tag{2.2}$$

A straightforward computation shows that (2.2) is equivalent to

$$\frac{2(1-p)^{\epsilon/2}\log_{\frac{1}{1-p}}(n)}{d(n) + 2\log_{\frac{1}{1-p}} 2 + \epsilon} \leq 1. \tag{2.3}$$

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 (2.3) is satisfied if

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

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

3. The Degree of Balance

In Theorem 3.3, formulated later in this section, we study relative m-balance 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 preparation for the proof of Theorem 3.3, we first prove two lemmas, which are of independent interest.

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

$$\{H_i \text{ is positive in } G_{n,p,p} \mid H \subseteq E(\tilde{G}_{n,p,p})\}, \quad \textit{for } i = 1, \ldots, t,$$

are mutually 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 $\{H_1, \ldots, H_t\}$. We have to prove that

$$P_H\left[H_i \text{ and } H_j \text{ are both positive}\right] = P_H\left[H_i \text{ is positive}\right] \times P_H\left[H_j \text{ is positive}\right].$$

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

$$P_H\left[H_i \text{ and } H_j \text{ are both positive}\right] = \frac{1}{4}.$$

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

$$\begin{aligned} &P_H\left[H_i \text{ and } H_j \text{ are both positive}\right] \\ &\quad = P_H\left[H_i \text{ is positive}\right] \times P_H\left[H_j \setminus H_i \text{ is positive}\right] = \frac{1}{4}. \end{aligned}$$

Consider now the case that $H_i \cap H_j \neq \varnothing$ 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

$$\begin{aligned} &P_H[H_i \text{ and } H_j \text{ are both positive}] \\ &\quad = P_H\left[H_i \setminus H_j \text{ is positive}\right] \times P_H\left[H_j \setminus H_i \text{ is positive}\right] \\ &\qquad \times P_H\left[H_i \cap H_j \text{ is positive}\right] \\ &\qquad + P_H\left[H_i \setminus H_j \text{ is negative}\right] \times P_H\left[H_j \setminus H_i \text{ is negative}\right] \\ &\qquad \times P_H\left[H_i \cap H_j \text{ is negative}\right]. \end{aligned}$$

By Lemma 2.1, each probability on the right-hand side of the above equality is equal to $\frac{1}{2}$. Thus

$$P_H\left[H_1 \text{ and } H_2 \text{ are both positive}\right] = \frac{1}{4}.$$

□

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 every arbitrarily small $\epsilon > 0$,*

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

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

$$E(X_m) \sim \frac{n^m}{2m}(2p)^m. \tag{3.1}$$

Variance of X_m:

$$\begin{aligned} E(X_m^2) &= E\left[\sum_{C_m \in \mathfrak{C}_m} 1_{\{G_{n,p,p} \supseteq C_m\}}\right]^2 \\ &= \sum_{C_m, C'_m \in \mathfrak{C}_m} \Pr\left[G_{n,p,p} \supseteq C_m \text{ and } G_{n,p,p} \supseteq C'_m\right] \\ &= \sum_{k=0}^{m}\left[\sum_{|C_m \cap C'_m|=k} \Pr\left[G_{n,p,p} \supseteq C_m \text{ and } G_{n,p,p} \supseteq C'_m\right]\right], \end{aligned} \tag{3.2}$$

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

For $k = 0$,

$$\begin{aligned} &\sum_{|C_m \cap C'_m|=0} \Pr\left[G_{n,p,p} \supseteq C_m \text{ and } G_{n,p,p} \supseteq C'_m\right] \\ &\quad = \left[\frac{(m-1)!}{2}\right]^2 \binom{n}{m}\binom{n-m}{m}(2p)^{2m} \sim \frac{n^{2m}(2p)^{2m}}{4m^2}. \end{aligned} \tag{3.3}$$

For $k \geq 1$,

$$\begin{aligned} &\sum_{|C_m \cap C'_m|=k} \Pr\left[G_{n,p,p} \supseteq C_m \text{ and } G_{n,p,p} \supseteq C'_m\right] \\ &\quad \leq (m!)^2\binom{n}{m}\binom{m}{k}\binom{n-m}{m-k}(2p)^{2m-k} \leq m^{2k}n^{2m-k}(2p)^{2m-k}. \end{aligned} \tag{3.4}$$

By (3.2), (3.3), and (3.4),

$$
\begin{aligned}
E(X_m^2) &\leq \frac{n^{2m}(2p)^{2m}}{4m^2} + \sum_{k=1}^{m} m^{2k} n^{2m-k} (2p)^{2m-k} \\
&\leq \frac{n^{2m}(2p)^{2m}}{4m^2} + \sum_{k=1}^{m} m^{2m} n^{2m-1} (2p)^{m} \\
&\leq \frac{n^{2m}(2p)^{2m}}{4m^2} + m^{2m+1} n^{2m-1} (2p)^{m}.
\end{aligned} \tag{3.5}
$$

From (3.1) and (3.5) 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, we obtain

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

which completes the proof. $\square$

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:

$$\rho(m) = \begin{cases} X_m^+/X_m & \text{if } X_m \neq 0, \\ 1/2 & \text{if } X_m = 0. \end{cases}$$

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

Proof. Let 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, $0 \leq k \leq \frac{(m-1)!}{2}\binom{n}{m}$. Let $\mathrm{C}_{m,k}$ be a fixed subset of C_m of cardinality k. We denote by $X_{m,k}^+$ the random variable X_m^+ conditioned by the event $\{\mathfrak{C}_m(\tilde{G}_{n,p,p}) = \mathfrak{C}_{m,k}\}$,

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

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

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

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

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

Thus

$$E(X_{m,k}) = \frac{k}{2}. \tag{3.6}$$

By Lemma 3.1, for $C_m,\ C'_m \in \mathfrak{C}_m,\ C_m \neq C'_m$, the events

$$\{C_m \text{ is positive} \mid \mathfrak{C}_m(\tilde{G}_{n,p,p}) = \mathfrak{C}_{m,k}\}$$

and

$$\{C'_m \text{ is positive} \mid \mathfrak{C}_m(\tilde{G}_{n,p,p}) = \mathfrak{C}_{m,k}\}$$

are independent. Thus

$$\operatorname{var}(X^+_{m,k}) = \frac{k}{4}. \tag{3.7}$$

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

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

where the sum is over all subsets $\mathfrak{C}_{m,k}$ of $\mathfrak{C}_m$. Using (3.6), we obtain

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

Since the above sum is equal to 1, we get

$$E(X^+_m \mid X_m = k) = \frac{k}{2}.$$

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

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

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

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

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

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

Equality (3.7) gives

$$\operatorname{var}\left[X_m^+ \mid X_m = k\right] = \frac{k}{4}.$$

Hence for $1 \leq k \leq \frac{(m-1)!}{2}\binom{n}{m}$, we have

$$\operatorname{var}\left[\rho(m) \mid X_m = k\right] = \frac{1}{4k},$$

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

$$\operatorname{var}\left[\rho(m) \mid X_m = 0\right] = 0.$$

Thus

$$\begin{aligned} \operatorname{var}[\rho(m)] &= \sum_{k=0}^{\frac{(m-1)!}{2}\binom{n}{m}} \operatorname{var}\left[\rho(m) \mid X_m = k\right] \times \Pr\left[X_m = k\right] \\ &= \sum_{k=1}^{\frac{(m-1)!}{2}\binom{n}{m}} \frac{1}{4k} \Pr\left[X_m = k\right]. \end{aligned} \tag{3.8}$$

Let ϵ be arbitrarily small positive real number. Then

$$\begin{aligned} &\Pr\left[X_m = k\right] \\ &\quad = \Pr\left[X_m = k \mid |X_m - E(X_m)| > \epsilon E(X_m)\right] \times \Pr\left[|X_m - E(X_m)| > \epsilon E(X_m)\right] \\ &\qquad + \Pr\left[X_m = k \mid |X_m - E(X_m)| \leq \epsilon E(X_m)\right] \\ &\qquad \times \Pr\left[|X_m - E(X_m)| \leq \epsilon E(X_m)\right]. \end{aligned}$$

Since $\frac{(m-1)!}{2}\binom{n}{m} \leq n^m$, by (3.8), we have

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

where the second summation is over

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

Lemma 3.2 gives

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

where the second summation is again over

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

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

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

Since ϵ is an arbitrarily small positive number, by setting $\epsilon = 1/\log n$ in this last inequality, we obtain

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

An application of Chebyshev's inequality completes the proof. □

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*

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

as $n \to \infty$.

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 $n(n-1)/2$ and p. Thus

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

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

$$\Pr\left[\left|Y_{S,T} - \frac{n^2p}{2}\right| > \frac{\epsilon n^2 p}{2}\right] \leq \exp\left[-\frac{\epsilon^2 n^2 p}{6}\right]. \tag{4.1}$$

In [EL Maftouhi 94], it was 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

$$\begin{aligned} &\Pr\left[\text{for all } \{S,T\},\ (1-\epsilon)\frac{n^2p}{2} \leq Y_{S,T} \leq (1+\epsilon)\frac{n^2p}{2}\right] \\ &\leq \Pr\left[(1-\epsilon)\frac{n^2p}{2} \leq \delta \leq (1+\epsilon)\frac{n^2p}{2}\right], \end{aligned}$$

it follows that

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

Using (4.1), we obtain

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

The right-hand side of the above inequality tends to 1 as $n \to \infty$. □

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

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

Theorem 5.1. *If $pn \to 0$, then $G_{n,p,p}$ is almost always balanced. If $p \geq c/n$, where $c > 2\log 2$ is a constant, then $G_{n,p,p}$ is almost always 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 $pn \to 0$, then $E(X) \to 0$ as well. From Markov's inequality we conclude that $G_{n,p,p}$ is almost always acyclic. Thus almost every signed graph is balanced.

Suppose now that $p > c/n$. With the notation introduced in the proof of Theorem 4.1, let

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

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

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

we have

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

Thus

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

The inequality $c > 2\log 2$ implies that $E(Z) = o(1)$, and by Markov's inequality, we conclude that $P[Z = 0] \to 1$ as $n \to \infty$. Thus the statement of the theorem follows from Theorem 1.1. □

References

[Bollobás 85] B. Bollobás. *Random Graphs.* Academic Press, 1985.

[Cartwright and Harary 56] D. Cartwright and F. Harary. "Structural Balance: A Generalization of Heider's Theory." *Psychological Review* 63 (1956), 277–293.

[Easley and Kleinberg 10] David Easley and Jon Kleinberg. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge University Press, 2010.

[EL Maftouhi 94] A. EL Maftouhi. "Probabalistic Methods on Combinatorics Graph Theory." PhD thesis, University of Paris XI (Orsay), 1994.

[Flament 65] G. Flament. *Théorie des Graphes et Structures Sociales*. Gauthier-Villars, Paris, 1965.

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

[Harary 54] F. Harary. "On the Notion of Balance of a Signed Graph." *Michigan Math. J.* 2 (1954), 143–146.

[Harary 59] F. Harary. "On the Measurement of Structural Balance." *Behavioral Sci.* 4 (1959), 316–323.

[Harary et al. 65] F. Harary, R. Z. Norman, and D. Cartwright. *Structural Models: An Introduction to the Theory of Directed Graphs*. Wiley, 1965.

[Heider 46] F. Heider. "Attitudes and Cognitive Organisation." *Journal of Psychology* 21 (1946), 107–112.

[Matula 76] D. W. Matula. "The Largest Clique Size in a Random Graph." Technical Report, Dept. of Computer Science, Southern Methodist University, Dallas, 1976.

[Morissete 58] J. Morissete. "On Experimental Study of the Theory of Structural Balance." *Human Relations* 11 (1958), 239–254.

[Norman and Roberts 78] 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 (1978), 66–91.

[Palmer 85] E. M. Palmer. *Graphical Evolution*. Wiley, 1985.

[Roberts 78] F. Roberts. *Graph Theory and Its Applications to Problems of Society*, CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM, 1978.

[Taylor 70] H. F. Taylor. *Balance in Small Groups*. Van Nostrand Reinhold, 1970.

A. EL Maftouhi, Laboratoire de Recherche en Informatique, Université Paris Sud 11, Bât. 650, 91405 Orsay Cedex, France (hakim.maftouhi@orange.fr)

Y. Manoussakis, Laboratoire de Recherche en Informatique, Université Paris Sud 11, Bât. 650, 91405 Orsay Cedex, France (yannis@lri.fr)

O. Megalakaki, Faculté de Philosophie, Université de Picardie, Chemin du Thil, 80025 Amiens Cedex 01, France (olga.megalakaki@u-picardie.fr)