Taylor & Francis
Taylor & Francis Group

WEAK BALANCE IN RANDOM SIGNED GRAPHS

Abdelhakim El Maftouhi, Ararat Harutyunyan, and Yannis Manoussakis

Laboratoire de Recherche en Informatique, Université Paris Sud 11, Orsay Cedex, France

Abstract *In this article, we strengthen some results in our previous work on balance in random signed graphs and study the weak balance. We show that many of the phenomena observed for the balance of random signed graphs extend to weak balance.*

1. INTRODUCTION AND TERMINOLOGY

In this work we deal with questions on balance and weak balance in random graphs. The theory of balance goes back to Heider [9] 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 [6, 13]. From a mathematical point of view, the most appropriate model for studying structural balance is that of signed graphs. Formally, a signed graph (G, σ) is a graph $G = (V, E)$ together with a function $\sigma : E \to \{+, -\}$, 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 it is said to be negative. A signed graph G is *balanced* if each cycle of G is positive, otherwise it is unbalanced. In 1956, Cartwright and Harary [3] obtained the following important result, which states an equivalent definition of balance.

Theorem 1.1. *A signed graph is balanced if and only if its vertex set can be partitioned into two classes so that every edge joining vertices within a class is positive and every edge joining vertices between classes is negative.*

One might also consider a weaker notion of balance. A signed graph is called *weakly balanced* if it contains no cycle with exactly one negative edge. Interestingly, for balanced graphs, we also have a nice characterization of weakly balanced graphs as Theorem 1.2 (see [4, 6]) shows.

Address correspondence to Ararat Harutyunyan, Laboratoire de Recherche en Informatique, Université Paris Sud 11, Bât. 650, 91405 Orsay Cedex, France. E-mail: ararat@lri.fr

Theorem 1.2. *A signed graph is weakly balanced if and only if its vertex set can be partitioned into different classes so that every edge connecting two vertices that belong to the same class is positive and every edge connecting two vertices that belong to different classes is negative.*

Balance in random social systems is considered in [7] (see also [8]). In [7], the authors defined a probabilistic model in which relations between individuals are assumed to be random. A good mathematical model for representing such random social structures is the so-called random signed graph $G(n, p, q)$ defined as follows. Let $p, q > 0$ be fixed, $0 < p+q < 1$. Given a set of n vertices, between each pair of distinct vertices there is either a positive edge with probability p or a negative edge with probability q, or else there is no edge with probability $1 - (p + q)$. The edges between different pairs of vertices are chosen independently. Throughout this article the expression "*asymptotically almost surely*" (a.a.s.) means "*with probability tending to* 1 *as n tends to infinity.*" In the random signed graph model introduced in [7], the authors proved that, a.a.s., $G(n, p, q)$ is unbalanced. Then they established bounds on the maximum order of a balanced induced subgraph. They also studied the relative m-balance and the frustration index (sometimes, also called the line index of balance; definitions provided following). Further recent studies have also looked at dynamic aspects of structural balance theory, modeling how the set of friendships and antagonisms in a graph, in other words, the labeling of the edges, might evolve over time as the social network implicitly seeks out structural balance. Antal, Krapivsky, and Redner [1] study a model that starts with a random labeling (choosing + or - randomly for each edge) and repeatedly balance each unbalanced triangle by flipping one of its labels. The model turns out to resemble the mathematical models one uses for certain physical systems as they reconfigure to minimize their energy [1, 11].

Throughout this article we will use the following definitions and notations. The *(weak) frustration index* δ of a signed graph G is the smallest number of edges whose removal from G results in a (weakly) balanced graph. We denote by X_m the total number of (positive and negative) cycles of length m in $G(n, p, q)$. By X_m^u we denote the number of *unbalanced* cycles of length m, i.e., the cycles of length m that contain exactly one negative edge. By X_m^- we denote the number of cycles of length m that have an odd number of negative edges. The *relative weak unbalance* is defined as the ratio $\rho(m) = X_m^u / X_m$. Similarly, the *relative (classical) unbalance* is defined as $\rho_c(m) = X_m^- / X_m$.

The rest of this article is organized as follows. In Section 2, we study the threshold for balance and weak balance in random graphs, including strengthening one of the main results in [7]. In Section 3, we study the order of the maximum weakly balanced component of $G_{n,p,q}$. In Section 4, we consider the weak frustration index. We conclude the article with a study of the relative weak unbalance in Section 5.

2. THRESHOLD FOR BALANCE AND WEAK BALANCE

El Maftouhi et al. [7] proved that the random signed graph $G(n, p, p)$, where $p = c/n$ for a sufficiently large constant c, is asymptotically almost surely unbalanced. Here, we strengthen this result by dropping the assumption that $p = q$. We first present some preliminary results that will be used later in this section.

2.1. Monotonicity

Recall that a property $\mathcal{Q}$ of graphs is *increasing* if it is preserved under the addition of edges. It is well known in the random graph model $G(n, p)$ that if $\mathcal{Q}$ is an increasing property, then $\mathbb{P}[G(n, p_1)$ *satisfies* $\mathcal{Q}] \leq \mathbb{P}[G(n, p_2)$ *satisfies* $\mathcal{Q}]$, whenever $p_1 \leq p_2$. We show that this natural inequality extends to random signed graphs.

A *signed graph property* is a set of signed graphs closed under (signed) isomorphism. We recall that two signed graphs (G_1, σ_1) and (G_2, σ_2) are isomorphic if there is a graph isomorphism θ that preserves the signs of the edges, that is, an isomorphism $\theta : G_1 \to G_2$ such that $\sigma_1(uv) = \sigma_2(\theta(uv))$ for each edge uv. A property of signed graphs is *increasing* if it is preserved under the addition of (signed) edges.

Lemma 2.1. *Let $\mathcal{Q}$ be an increasing property of signed graphs. Let $0 \leq p_1 \leq p_2 \leq 1$, $0 \leq q_1 \leq q_2 \leq 1$ with $0 \leq p_1 + q_1 \leq 1$ and $0 \leq p_2 + q_2 \leq 1$. Then*

$$\mathbb{P}\big[G(n, p_1, q_1) \text{ satisfies } \mathcal{Q}\big] \leq \mathbb{P}\big[G(n, p_2, q_2) \text{ satisfies } \mathcal{Q}\big].$$

Proof. To prove this lemma we will use a version of the two-round exposure technique adapted to random signed graphs. Let $G(n, p_0, q_0)$ be a random signed graph generated independently on the same vertex set as $G(n, p_1, q_1)$. Let $G(n, p', q')$ be the random signed graph obtained by adding to $G(n, p_1, q_1)$ all the edges of $G(n, p_0, q_0)$ that do not belong to $G(n, p_1, q_1)$ (with their signs), that is,

$$E(G(n, p', q')) = E(G(n, p_1, q_1)) \cup [E(G(n, p_0, q_0)) \setminus E(G(n, p_1, q_1))].$$

Thus

$$p' = p_1 + (1 - p_1 - q_1)p_0 \ \text{ and } \ q' = q_1 + (1 - p_1 - q_1)q_0.$$

By setting

$$p_0 = \frac{p_2 - p_1}{1 - p_1 - q_1} \ \text{ and } \ q_0 = \frac{q_2 - q_1}{1 - p_1 - q_1},$$

it follows that $G(n, p_2, q_2)$ and $G(n, p', q')$ have the same probability distribution. To the contrary, as $G(n, p_1, q_1) \subset G(n, p', q')$, and $\mathcal{Q}$ is an increasing property, the event $\{G(n, p_1, q_1)$ *satisfies* $\mathcal{Q}\}$ is included in the event $\big\{G(n, p', q')$ *satisfies* $\mathcal{Q}\big\}$, which completes the proof.   □

2.2. The Uniform Model and Asymptotic Equivalence

Given $1 \leq M \leq \binom{n}{2}$ and $0 \leq \alpha \leq 1$, the *uniform random signed graph* on n vertices, denoted by $G(n, M, \alpha)$, is obtained by choosing, at random, M edges and assigning independently to each edge a positive sign with probability α or a negative sign with probability $1 - \alpha$. Formally, for every fixed signed graph G of size M with $|E^+(G)|$ positive edges and $|E^-(G)|$ negative edges, we have

$$\mathbb{P}\left[G(n, M, \alpha) = G\right] = \frac{\alpha^{|E^+(G)|}(1 - \alpha)^{|E^-(G)|}}{\binom{\binom{n}{2}}{M}}.$$

It is not difficult to see that the uniform random signed graph $G(n, M, \alpha)$ can be viewed as the Mth stage of the random signed graph process $\{G(n, M, \alpha)\}_M, 0 \leq M \leq \binom{n}{2}$, defined as follows. We start with no edges at time 0, and at each step a new edge is selected uniformly at random from edges not already selected and is assigned a positive sign with probability α or a negative sign with probability $1 - \alpha$.

The following result shows in particular that, under certain conditions, if a property $\mathcal{Q}$ holds a.a.s. for the model $G(n, M, \alpha)$, it also holds a.a.s. for the model $G(n, p, q)$, for appropriate values of p and q. The result does not require any restriction on $\mathcal{Q}$.

Theorem 2.2. *Let $\mathcal{Q}$ be a property of signed graphs. Let $p = p(n) \geq 0$, $q = q(n) \geq 0$ with $0 \leq p + q \leq 1$. Let $0 \leq a \leq 1$. If for every $M = M(n)$ such that*

$$M = \binom{n}{2}(p + q) + O\left(\sqrt{\binom{n}{2}(p + q)(1 - p - q)}\right),$$

we have $\mathbb{P}\left[G(n, M, p/(p + q)) \text{ satisfies } \mathcal{Q}\right] \rightarrow a$ as $n \rightarrow \infty$, then also we have $\mathbb{P}\left[G(n, p, q) \text{ satisfies } \mathcal{Q}\right] \rightarrow a$ as $n \rightarrow \infty$.

We omit the proof of this theorem because the arguments used are similar to those employed in classical random graphs; see, for example, [10]. The proof uses the elementary law of total probability and Chebyshev's inequality. We just need to check that

$$\mathbb{P}\left[G(n, p, q) \text{ satisfies } \mathcal{Q} \,\big|\, |E(G(n, p, q))| = M\right] = \mathbb{P}\left[G(n, M, p/(p + q)) \text{ satisfies } \mathcal{Q}\right],$$

which is straightforward.

2.3. Threshold for Balance

It is well known that the threshold for the existence of cycles in the random graph model $G(n, p)$ is $1/n$. Thus, if $p + q = o(1/n)$ then $G(n, p, q)$ is a.a.s. balanced. The following theorem gives conditions on p and q under which $G(n, p, q)$ is a.a.s. unbalanced.

Theorem 2.3. *Let $\epsilon, \delta > 0$ be fixed. If $p + q \geq (1 + \epsilon)/n$ and $q \geq \delta/n$ then $G(n, p, q)$ is a.a.s. unbalanced.*

Proof. Because *"unbalance"* is an increasing property, by Lemma 2.1, it is sufficient to prove the theorem for $p + q = (1 + \epsilon)/n$ and $q = \delta/n$. As defined before, let $\{G(n, M, \alpha)\}_M$, where $\alpha = p/(p + q) = (1 + \epsilon - \delta)/(1 + \epsilon)$, be the associated random graph process. By Theorem 2.2, it suffices to show that $G(n, M, \alpha)$ is a.a.s. unbalanced when $M = (1 + \epsilon)n/2$. From the well-known result of Erdős and Renyi concerning the phase transition phenomenon of random graphs, we know that at stage $M_0 = (1 + \epsilon/2)n/2$, $G(n, M_0, \alpha)$, a.a.s. has a giant component of size at least $c(\epsilon)n$, where $c(\epsilon)$ is a constant depending only on ϵ. Let V_0 and $E(V_0)$ be the vertex set and the edge set of the giant component at time M_0, respectively. So we have

$$\mathbb{P}[|V_0| \geq c(\epsilon)n] \geq 1 - o(1). \tag{2.1}$$

After this stage, we add the remaining $\epsilon n/4$ edges one by one. For every $1 \le i \le \epsilon n/4$, let $e_i = x_i y_i$ denote the edge added to the graph at time $M_i = M_0 + i$. If x_i and y_i are both in V_0, then there exists a path $P_{x_i y_i}$ joining x_i and y_i. Thus,

$$
\begin{aligned}
\mathbb{P}\big[e_i \text{ closes a negative cycle} \mid x_i, y_i \in V_0\big] \ge \ & \mathbb{P}\big[P_{x_i y_i} \text{ is positive, } e_i \text{ is} \\
& \text{negative} \mid x_i, y_i \in V_0\big] \\
& + \mathbb{P}\big[P_{x_i y_i} \text{ is negative, } e_i \text{ is} \\
& \text{positive} \mid x_i, y_i \in V_0\big] \\
\ge \ & \min(\alpha, 1 - \alpha).
\end{aligned}
$$

The probability that an edge e_i has its end vertices in V_0 is

$$
\mathbb{P}\big[\, x_i, y_i \in V_0 \big] = \frac{\binom{|V_0|}{2} - |[V_0]^2 \cap E(G(n, M_{i-1}, \alpha))|}{\binom{n}{2} - M_0 - i + 1}.
$$

Therefore,

$$
\begin{aligned}
& \mathbb{P}\big[e_i \text{ closes a negative cycle in } E(V_0)\big] \\
& \qquad \ge \frac{\binom{|V_0|}{2} - \big|[V_0]^2 \cap E(G(n, M_{i-1}, \alpha))\big|}{\binom{n}{2} - M_0 - i + 1} \min(\alpha, 1 - \alpha).
\end{aligned}
$$

Since

$$
\begin{aligned}
\mathbb{P}\big[e_i \text{ closes a negative cycle}\big] \ge \ & \mathbb{P}\big[e_i \text{ closes a negative cycle in } E(V_0) \mid |V_0| \\
& \ge c(\epsilon)\big]\mathbb{P}\big[|V_0| \ge c(\epsilon)\big],
\end{aligned}
$$

and by (2.1), we have

$$
\begin{aligned}
\mathbb{P}\big[e_i \text{ closes a negative cycle}\big] \ge \ & (1 - o(1))\frac{\binom{c(\epsilon)n}{2} - (1 + \frac{\epsilon}{2})\frac{n}{2} - i + 1}{\binom{n}{2} - (1 + \frac{\epsilon}{2})\frac{n}{2} - i + 1} \min(\alpha, 1 - \alpha) \\
\ge \ & c^2(\epsilon) \min(\alpha, 1 - \alpha)(1 - o(1)).
\end{aligned}
$$

Thus,

$$
\mathbb{P}\big[e_i \text{ closes a negative cycle}\big] \ge C(\epsilon, \delta),
$$

where $C(\epsilon, \delta) < 1$ is a constant (not depending on n). It follows that

$$
\begin{aligned}
& \mathbb{P}\big[\text{for some } 1 \le i \le \epsilon n/4, \ \ e_i \text{ closes a negative cycle}\big] \\
& \qquad \ge 1 - (1 - C(\epsilon, \delta))^{\frac{\epsilon n}{4}} = 1 - o(1),
\end{aligned}
$$

and thus $G(n, p, q)$ is unbalanced, a.a.s. $\qquad\square$

2.4. Threshold for Weak Balance

Using the ideas of the previous theorem, we can derive an analogous result for weak balance.

Theorem 2.4. *Let $\epsilon, \delta > 0$ be fixed. If $p \geq (1 + \epsilon)/n$ and $q \geq \delta/n$ then $G(n, p, q)$ is a.a.s. weakly unbalanced.*

Proof. We may assume that $p = (1 + \epsilon)/n$ an $q = \delta/n$. As defined in the previous theorem, let $\{G(n, M, \alpha)\}_M$, where $\alpha = p/(p+q) = (1+\epsilon)/(1+\epsilon+\delta)$, be the associated random graph process. At stage $M_0 = (1 + \epsilon/2)n/2$, the random graph with only positive edges a.a.s. has a giant component of order at least $c(\epsilon)n$, and, clearly, any edge e_i added at time $M_i = M_0 + i$ will close an unbalanced cycle (cycle with exactly one negative edge) if e_i is negative and if it has both end vertices in the giant component. Using a similar computation as in Theorem 2.3, one can easily show that, at time M_i, we have

$$\mathbb{P}\big[e_i \text{ closes an unbalanced cycle}\big] \geq \frac{c^2(\epsilon)\delta}{(1 + \epsilon + \delta)}(1 - o(1)).$$

Now, employing an argument similar to that in Theorem 2.3, we conclude that $G(n, p, q)$ a.a.s. is weakly unbalanced. $\square$

3. THE MAXIMUM ORDER OF A WEAKLY BALANCED SUBGRAPH

Let $G(n, p, q)$ be the random signed graph on n vertices where p and q are fixed and $0 < p + q < 1$. It follows from the last theorem of the previous section that $G(n, p, q)$ is weakly unbalanced. We then provide estimates of the maximum size of a weakly balanced induced subgraph in $G(n, p, q)$.

Let $\beta' = \beta'(G(n, p, q))$ denote the maximum order of a weakly balanced induced subgraph of $G(n, p, q)$. Because each induced subgraph of $G(n, p, q)$ containing no negative edges or no positive edges is weakly balanced, we have

$$\beta'(G(n, p, q)) \geq \alpha(G(n, p)) \text{ and } \beta'(G(n, p, q)) \geq \alpha(G(n, q)), \qquad (3.1)$$

where $\alpha(G(n, p))$ denotes the independence number of the random graph $G(n, p)$.

It is well known that the independence number of $G(n, p)$ is a.a.s. concentrated on two consecutive integer values, see [12]. More precisely, for any $\epsilon > 0$, we have

$$\Pr\Big[\lfloor d(n, p) - \epsilon \rfloor \leq \alpha(G(n, p)) \leq \lfloor d(n, p) + \epsilon \rfloor\Big] \to 1 \text{ as } n \to \infty, \qquad (3.2)$$

where

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

We will also use the following asymptotic formula of the Bell number B_n that appears in [5]. Recall that B_n is the number of ways to partition a set of n elements into different subsets.

$$\frac{\log(B_n)}{n} = \log(n) - \log(\log(n)) - 1 + o(1). \qquad (3.3)$$

Theorem 3.1. *Let $G(n, p, q)$ be the random signed graph on n vertices where p and q are fixed, $0 < p + q < 1$.*

If $p \geq q$ then, a.a.s., we have

$$2 \log_a n - 2 \log_a \log_a n + A \leq \beta'(G(n, p, q)) \leq 2 \log_a n - 2 \log_a \log_a \log_a n + B,$$

where $a = 1/(1 - q)$ and A and B are constants.
   If $p < q$ then, a.a.s., we have

$$2 \log_b n - 2 \log_b \log_b n + A' \leq \beta'(G(n, p, q)) \leq 2 \log_b n - 2 \log_b \log_b \log_b n + B',$$

where $b = 1/(1 - p)$ and A' and B' are constants.

Proof. For a given subset S of r vertices, we have

$$\mathbb{P}[S \text{ is weakly balanced }] = \sum_{k=1}^{r} \sum_{\{S_1, \dots, S_k\}} (1 - q)^{\sum_{i=1}^{k} \binom{n_i}{2}} (1 - p)^{\sum_{1 \leq i < j \leq k} n_i n_j},$$

where the second sum is over all partitions $\{S_1, \dots, S_k\}$ of S into k different subsets and $|S_i| = n_i$ for $1 \leq i \leq k$. Let N_r be the number of subsets of vertices of size r that induce a weakly balanced subgraph in $G(n, p, q)$.

   Case where $p \geq q$. The lower bound follows from (3.2) and the second inequality of (3.1). To prove the upper bound, it is sufficient to show, by using Markov inequality, that the expectation $\mathbb{E}(N_r)$ of N_r tends to 0 as $n \to \infty$. Now, using the fact that $1 - p \leq 1 - q$, we have that

$$\mathbb{E}(N_r) \leq \binom{n}{r} \mathbb{P}[S \text{ is weakly balanced}] \leq \binom{n}{r} B_r (1 - q)^{\frac{r(r-1)}{2}}.$$

Using Stirling's formula, we get

$$\mathbb{E}(N_r) \leq \frac{1}{\sqrt{2\pi r}} \left(\frac{en B_r^{\frac{1}{r}} (1 - q)^{\frac{r-1}{2}}}{r} \right)^r.$$

Thus, $\mathbb{E}(N_r) \to 0$ if

$$\frac{en B_r^{\frac{1}{r}} (1 - q)^{\frac{r-1}{2}}}{r} \leq 1. \tag{3.4}$$

Condition (3.4) is satisfied if

$$1 + \log n + \frac{\log B_r}{r} + \frac{r - 1}{2} \log(1 - q) - \log r \leq 0.$$

Now, set

$$r := 2 \log_a n - 2 \log_a \log_a \log_a n + B,$$

where B is a sufficiently large constant.
   By using (3.3), we now see that condition (3.4) is satisfied, as required.

Case where $p < q$. The lower bound follows from (3.2) and the first inequality of (3.1). For the upper bound, note that $1 - p > 1 - q$, from which it follows that

$$\mathbb{E}(N_r) \leq \binom{n}{r} \mathbb{P}[S \text{ is weakly balanced}] \leq \binom{n}{r} B_r (1 - p)^{\frac{r(r-1)}{2}}.$$

Now, an argument identical to that in the previous case finishes the proof. □

4. WEAK FRUSTRATION INDEX

Our aim is to prove the following theorem on the weak frustration index, an analog of the result found in [7] in the case of the frustration index. Although one could expect that the weak frustration index should be smaller than the frustration index, surprisingly, this is not the case.

Given a signed graph G, and a partition $\{S_1, \ldots, S_k\}$ of $V(G)$, define $Y_{S_1,\ldots,S_k} = \sum \Delta_{S_i}^- + \sum_{i<j} \Delta_{S_i,S_j}^+$, where $\Delta_{S_i}^-$ is the number of negative edges in the subgraph $G[S_i]$ induced by S_i, and Δ_{S_i,S_j}^+ is the number of positive edges that have one endpoint in S_i and one endpoint in S_j. We will use the following lemma whose proof is immediate.

Lemma 4.1. *For every signed graph G, $\delta(G) = \min_{P=\{S_1,\ldots,S_k\}} Y_{S_1,\ldots,S_k}$, where the minimum is taken over all partitions P of $V(G)$.*

Theorem 4.2. *Let $0 < p < 1/2$ be fixed. Then, for any fixed $\epsilon > 0$, the weak frustration index δ of $G(n, p, p)$ satisfies*

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

as $n \to \infty$.

Proof. Let $\{S_1, \ldots, S_k\}$ be a fixed partition of the vertex set of $G(n, p, p)$. Let $Y_{S_1,\ldots,S_k}$ be the random variable defined as before. It is easily seen that $Y_{S_1,\ldots,S_k}$ has a binomial distribution, $Y_{S_1,\ldots,S_k} \sim \text{Bin}\left(\binom{n}{2}, p\right)$. Thus, $\mathbb{E}[Y_{S_1,\ldots,S_k}] = n(n-1)p/2$. Since $\delta \leq Y_{S_1,\ldots,S_k}$, the inequality from above is an immediate consequence of the law of large numbers. To prove the lower bound, we use the following (upper tail) Chernoff bound (see [2], page 12).

$$\mathbb{P}\left[Y_{S_1,\ldots,S_k} < (1 - \epsilon)\frac{n^2 p}{2}\right] \leq \exp\left[\frac{\epsilon^2 n^2 p}{4}\right]. \tag{4.1}$$

Now, by Lemma 4.1, we have

$$\mathbb{P}\left[\delta < (1 - \epsilon)\frac{n^2 p}{2}\right] \leq \mathbb{P}\left[\exists \text{ a partition } \{S_1, \ldots, S_k\}, \ Y_{S_1,\ldots,S_k} < (1 - \epsilon)\frac{n^2 p}{2}\right]$$

$$\leq B_n \mathbb{P}\left[Y_{S_1,\ldots,S_k} < (1 - \epsilon)\frac{n^2 p}{2}\right],$$

where B_n denotes the Bell number. Relations (3.3) and (4.1) imply

$$\mathbb{P}\left[\delta < (1 - \epsilon)\frac{n^2 p}{2}\right] \leq n^n \exp\left[-\frac{\epsilon^2 n^2 p}{4}\right] = o(1).$$

□

5. RELATIVE WEAK UNBALANCE

In this section we study relative weak unbalance by examining the ratio $\rho(m) = X_m^u / X_m$ of the number X_m^u of unbalanced m-cycles to the total number X_m of m-cycles in $G(n, p, q)$. We set $\rho(m) = 0$ if there are no cycles of length m in $G(n, p, q)$. Recall that a cycle is *unbalanced* if it contains exactly one negative edge. Clearly, the expectation of X_m is given by

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

In order to prove Theorem 5.2, we use the following lemma from [7] concerning the concentration of X_m around its expected value.

Lemma 5.1. *Let m be a fixed integer, $3 \leq m \leq n$. Then, for any arbitrarily small $\epsilon > 0$,*

$$\mathbb{P}\Big[\, |\, X_m - \mathbb{E}(X_m)\,| \;>\; \epsilon\mathbb{E}(X_m)\,\Big] \leq \frac{4m^{2m+3}}{\epsilon^2 n(p+q)^m}. \tag{5.1}$$

Theorem 5.2. *Let m be a fixed integer, $m \geq 3$. Let $p, q > 0$ be fixed, $0 < p + q \leq 1$. Then, for any $\epsilon > 0$, the relative weak unbalance $\rho(m)$ of $G(n, p, q)$ satisfies*

$$\mathbb{P}\Big[\big|\rho(m) - \frac{qp^{m-1}}{(p+q)^m}\big| \leq \epsilon\Big] \to 1 \quad \text{as } n \to \infty.$$

Proof. We denote by $\mathcal{C}_m = \mathcal{C}_m(G(n, p, q))$ the set of all m-cycles of $G(n, p, q)$. Let C be a fixed m-cycle. Clearly, we have

$$\mathbb{P}\Big[C \text{ is unbalanced} \,\Big|\, C \in \mathcal{C}_m(G(n, p, q))\Big] = \frac{mqp^{m-1}}{(p+q)^m}.$$

Expectation of ρ_m. For $k \geq 1$, we have

$$\mathbb{E}\Big(X_m^u \,\Big|\, X_m = k\Big) = k\frac{mqp^{m-1}}{(p+q)^m}.$$

Thus,

$$\mathbb{E}\Big(\rho(m) \,\Big|\, X_m = k\Big) = \frac{mqp^{m-1}}{(p+q)^m},$$

and so

$$\mathbb{E}\Big(\rho(m)\Big) = \frac{mqp^{m-1}}{(p+q)^m}.$$

Variance of $\rho(m)$. We need to prove that, for any $\epsilon > 0$,

$$\mathbb{P}\Big[|\rho(m) - \mathbb{E}(\rho(m))| > \epsilon\Big] \to 0 \text{ as } n \to \infty.$$

By conditioning on the event $\{|X_m - \mathbb{E}(X_m)| > \epsilon\mathbb{E}(X_m)\}$ and its complement, we get

$$\mathbb{P}\Big[|\rho(m) - \mathbb{E}(\rho(m))| > \epsilon\Big] \le \mathbb{P}\Big[|X_m - \mathbb{E}(X_m)| > \epsilon\mathbb{E}(X_m)\Big] +$$

$$\mathbb{P}\Big[|\rho(m) - \mathbb{E}(\rho(m))| > \epsilon \wedge (1-\epsilon)\mathbb{E}(X_m) \le X_m \le (1+\epsilon)\mathbb{E}(X_m)\Big]. \quad (5.2)$$

By Lemma 5.1, the first term of the right-hand side of the above inequality tends to 0 whenever $\epsilon\sqrt{n} \to \infty$ as $n \to \infty$. Then, it suffices to show that the second term tends to 0 as $n \to \infty$.

Clearly

$$\mathbb{P}\Big[|\rho(m) - \mathbb{E}(\rho(m))| > \epsilon \wedge (1-\epsilon)\mathbb{E}(X_m) \le X_m \le (1+\epsilon)\mathbb{E}(X_m)\Big]$$

$$= \sum_{(1-\epsilon)\mathbb{E}(X_m) \le k \le (1+\epsilon)\mathbb{E}(X_m)} \mathbb{P}\Big[|\rho(m) - \mathbb{E}(\rho(m))| > \epsilon \wedge X_m = k\Big]$$

$$= \sum_k \sum_{C_1,\ldots,C_k} \mathbb{P}\Big[|\rho(m) - \mathbb{E}(\rho(m))| > \epsilon \,\Big|\, \mathscr{C}_m(G(n, p, q)) = \{C_1, \ldots, C_k\}\Big]$$

$$\times \mathbb{P}\Big[\mathscr{C}_m(G(n, p, q)) = \{C_1, \ldots, C_k\}\Big],$$

where the second sum is over all possible sets of k cycles of size m in $G(n, p, q)$. In order to apply properly Chebyshev's inequality, we observe that for any fixed set $\{C_1, \ldots, C_k\}$ of k m-cycles, we have

$$\mathbb{E}(\rho(m)) = \mathbb{E}\left(\rho(m) \,\Big|\, \mathcal{C}_m(G(n, p, q)) = \{C_1, \ldots, C_k\}\right).$$

Thus,

$$\mathbb{P}\Big[|\rho(m) - \mathbb{E}(\rho(m))| > \epsilon \wedge (1-\epsilon)\mathbb{E}(X_m) \le X_m \le (1+\epsilon)\mathbb{E}(X_m)\Big]$$

$$\le \frac{1}{\epsilon^2} \sum_k \sum_{C_1,\ldots,C_k} \text{var}\Big[\rho(m) \,\Big|\, \mathscr{C}_m(G(n, p, q)) = \{C_1, \ldots, C_k\}\Big]$$

$$\times \mathbb{P}\Big[\mathscr{C}_m(G(n, p, q)) = \{C_1, \ldots, C_k\}\Big]. \quad (5.3)$$

Let $\{C_1, \ldots, C_k\}$ be a fixed set of k m-cycles.

$$\mathbb{E}\Big[\rho^2(m) \,\Big|\, \mathscr{C}_m(G(n, p, q)) = \{C_1, \ldots, C_k\}\Big]$$

$$= \frac{1}{k^2}\mathbb{E}\Big[(X_m^u)^2 \,\Big|\, \mathscr{C}_m(G(n, p, q)) = \{C_1, \ldots, C_k\}\Big]$$

$$= \frac{1}{k^2} \sum_{1 \le i, j \le k} \mathbb{P}\Big[C_i, C_j \text{ are unbalanced} \,\Big|\, \mathscr{C}_m(G(n, p, q)) = \{C_1, \ldots, C_k\}\Big]$$

$$= \frac{1}{k^2} \sum_{C_i \cap C_j = \emptyset} \mathbb{P}\Big[C_i, C_j \text{ are unbalanced} \,\Big|\, \mathcal{C}_m(G(n, p, q)) = \{C_1, \ldots, C_k\}\Big]$$

$$+ \frac{1}{k^2} \sum_{C_i \cap C_j \neq \emptyset} \mathbb{P}\Big[C_i, C_j \text{ are unbalanced} \Big| \mathcal{C}_m(G(n, p, q)) = \{C_1, \ldots, C_k\}\Big]$$

$$\leq \left[\frac{1}{k} \sum_{1 \leq i \leq k} \mathbb{P}\Big[C_i \text{ is unbalanced} \Big| \mathcal{C}_m(G(n, p, q)) = \{C_1, \ldots, C_k\}\Big]\right]^2$$

$$+ \frac{1}{k^2} \sum_{l=1}^{m} \left(\sum_{|C_i \cap C_j|=l} \mathbb{P}\Big[C_i, C_j \text{ are unbalanced} \Big| \mathcal{C}_m(G(n, p, q))\right.$$

$$= \{C_1, \ldots, C_k\}\Big]\bigg),$$

where $|C_i \cap C_j| = l$ means that C_i and C_j have l vertices in common. For every $l \geq 1$, the number of pairs (C_i, C_j) having exactly l vertices in common is at most

$$((m-1)!/2)^2 \binom{n}{m} \binom{n}{m-l} = O(n^{2m-l}).$$

Because m is fixed (not depending on n), the last sum in the double sum above is at most $O(n^{2m-1})$. Thus,

$$\mathbb{E}\Big[\rho^2(m) \Big| \mathcal{C}_m(G(n, p, q)) = \{C_1, \ldots, C_k\}\Big]$$

$$\leq \mathbb{E}^2\Big[\rho(m) \Big| \mathcal{C}_m(G(n, p, q)) = \{C_1, \ldots, C_k\}\Big] + \frac{1}{k^2} O(n^{2m-1}).$$

That is,

$$\text{var}\Big[\rho(m) \Big| \mathcal{C}_m(G(n, p, q)) = \{C_1, \ldots, C_k\}\Big] \leq \frac{1}{k^2} O(n^{2m-1}).$$

Since $k \geq (1 - \epsilon)\mathbb{E}(X_m) = (1 - \epsilon)\Omega(n^m)$, we get

$$\text{var}\Big[\rho(m) \Big| \mathcal{C}_m(G(n, p, q)) = \{C_1, \ldots, C_k\}\Big] \leq \frac{1}{(1 - \epsilon)^2} O\left(\frac{1}{n}\right). \tag{5.4}$$

Now, (5.4), (5.3), (5.2) and (5.1) give

$$\mathbb{P}\Big[|\rho(m) - \mathbb{E}(\rho(m))| > \epsilon\Big]$$

$$\leq O\left(\frac{1}{n\epsilon^2}\right) + O\left(\frac{1}{n\epsilon^2(1-\epsilon)^2}\right) \sum_k \sum_{C_1,\ldots,C_k} \mathbb{P}\Big[\mathcal{C}_m(G(n, p, q)) = \{C_1, \ldots, C_k\}\Big]$$

$$\leq O\left(\frac{1}{n\epsilon^2}\right) + O\left(\frac{1}{n\epsilon^2(1-\epsilon)^2}\right) \sum_k \mathbb{P}\Big[X_m = k\Big]$$

$$\leq O\left(\frac{1}{n\epsilon^2}\right).$$

Thus, by setting $\epsilon = 1/\log n$, we conclude the proof. $\qquad\square$

An analog of Theorem 5.2 for balance was proved in [7]: the authors showed that a.a.s. in $G(n, p, p)$, $\rho_c(m) \to 1/2$ as $n \to \infty$. The method of the proof of Theorem 5.2 allows us to significantly strengthen this result as follows.

Theorem 5.3. *Let m be a fixed integer, $m \geq 3$. Suppose that p, q are fixed. Then, for any $\epsilon > 0$, the relative unbalance $\rho_c(m)$ of $G(n, p, q)$ satisfies*

$$\mathbb{P}\left[\left|\rho_c(m) - \frac{q}{p + q}\right| \leq \epsilon\right] \to 1 \quad as \ \ n \to \infty.$$

Because the proof of Theorem 5.3 is nearly identical to that of Theorem 5.2, we omit it.

ACKNOWLEDGMENT

FUNDING

REFERENCES

[1] T. Antal, P. Krapivsky, and S. Redner. "Social Balance on Networks: The Dynamics of Friendship and Enmity." *Physica D* 224(2006), 130–136.

[2] B. Bollobás. *Random Graphs,* Second Edition. Cambridge, UK: Cambridge University Press, 2001.

[3] D. Cartwright and F. Harary. "Structural Balance: A Generalization of Heider's Theory." *Psychological Review* 63:5(1956), 277–293.

[4] J. A. Davis. "Clustering and Structural Balance in Graphs." *Human Relations* 20:2(1967), 181–187.

[5] N. G. de Bruijn. *Asymptotic Methods in Analysis*. New York, NY: Dover, 1981.

[6] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. New York, NY: Cambridge University Press, 2010.

[7] A. El Maftouhi, Y. Manoussakis, and O. Megalakaki. "Balance in Random Signed Graphs." *Internet Mathematics* 8:4(2012), 364–380.

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

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

[10] S. Janson, T. Luczak, and A. Rucinski. *Random Graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. New York, NY: Wiley-Interscience, 2000.

[11] S. A. Marvel, S. H. Strogatz, and J. M. Kleinberg. "The Energy Landscape of Social Balance." *Physical Review Letters* 103:19(2009), 198701.

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

[13] T. Zaslavsky. "A Mathematical Bibliography of Signed and Gain Graphs and Allied Areas." *Electronic Journal of Combinatorics* 5, Dynamic Surveys 8(1998).