AUSTRALASIAN JOURNAL OF COMBINATORICS
Volume **75(1)** (2019), Pages 50–65

Balance in random signed intersection graphs

Konstantinos Papalamprou

Department of Electrical and Computer Engineering
Aristotle University of Thessaloniki
Thessaloniki, Greece
papalamprou@ece.auth.gr

Christoforos Raptopoulos

Computer Technology Institute and University of Patras
Patras, Greece
raptopox@ceid.upatras.gr

Abstract

We propose a new model for random signed graphs, namely the *signed random intersection graphs model*. In particular, each vertex of a signed graph is associated with a (finite) universal set of features. For each feature, every vertex can have either a positive, a negative or an indifferent *view* towards that feature with probability p, q and $1-p-q$ respectively. Based on the value of a metric that measures the level of agreement between a pair of vertices towards the set of predefined features, an edge may be added between these two vertices which has either a positive or a negative sign. Under this framework, we initiate the study of random signed intersection graphs by providing several preliminary results concerning the number of conflicting views among all vertices using the well-known notion of balance in signed graphs.

1 Introduction

In this paper, we consider a network setting where a connection between two members of the network can be either positive, negative or indifferent. These networks can be abstracted using *signed graphs*, in which there is a set of vertices and existing edges between pairs of vertices are labeled with either a plus (+) or a minus (−) sign. Since their introduction by Cartway and Harary [4, 5] in mid-fifties, signed graphs have been extensively studied in the literature (see the annotated bibliography [14] with regard to signed graphs and allied areas).

In a recent line of research, Maftouhi, Manoussakis and Megalakaki [8] and Maftouhi, Harutyunyan and Manoussakis [7] extended Cartwright-Harary's theory

ISSN: 2202-3518

of balance in deterministic social structures, by introducing *random signed graphs* $\mathcal{G}_{n,p,q}$. In their model, there are n vertices and each edge independently is either positive with probability p, or negative with probability q, or does not exist with probability $1 - p - q$. Equivalently, a random instance of $\mathcal{G}_{n,p,q}$ is constructed by taking an Erdős-Rényi random graph $G_{n,p+q}$ and independently labeling an existing edge with a $+$ sign with probability $\frac{p}{p+q}$ and with a $-$ sign with probability $\frac{q}{p+q}$.

Inspired by the work of Maftouhi et al. but also from feature-based construction of edges in random intersection graphs (see [6] and [10] for mathematical and algorithmic aspects, but also the excellent survey [3] where it is also discussed how such graphs capture many characteristics of social and complex networks), we propose here a new model for random signed graphs, namely the *signed random intersection graphs model*. Note that, in what follows, we denote graph instances by G and the corresponding probability distributions by $\mathcal{G}$. A formal definition of our model follows:

Definition 1 (Random Signed Intersection Graphs)**.** Let $\mathcal{V} = \{v_1, \dots, v_n\}$ be a set of n vertices and let $\mathcal{M} = \{\ell_1, \dots, \ell_m\}$ be a set of m *features*. We get a random instance of the *random signed intersection graphs model* $\mathcal{G}_{n,m,p,q}$ as follows: To each vertex v we assign two sets S_v^+ and S_v^- by independently choosing for each feature ℓ to either belong to S_v^+ with probability p, or to belong to S_v^- with probability q, or to belong to none of the two with probability $1 - p - q$. For each vertex v, let $\mathbf{x}_v \in \{-1, 0, +1\}^m$ be the vector given by $\mathbf{x}_v[\ell] = -1$ if $\ell \in S_v^-$, $\mathbf{x}_v[\ell] = +1$ if $\ell \in S_v^+$, and $\mathbf{x}_v[\ell] = 0$ otherwise. Two vertices u, v are connected with a *positive edge* if and only if $\sum_\ell \mathbf{x}_v[\ell]\mathbf{x}_u[\ell] > 0$, with a *negative edge* if $\sum_\ell \mathbf{x}_v[\ell]\mathbf{x}_u[\ell] < 0$ and not connected if $\sum_\ell \mathbf{x}_v[\ell]\mathbf{x}_u[\ell] = 0$.

In the above model, we denote by E^+ (respectively, E^-) the set of positive (respectively, negative) edges of $G_{n,m,p,q}$. Furthermore, we define the *associated signed bipartite graph* $B_{n,m,p,q}$ as the bipartite graph having vertex set $\mathcal{V} \cup \mathcal{M}$ and edge set $E_B^+ \cup E_B^-$, where $E_B^+ = \{(v, \ell) : \ell \in S_v^+\}$ and $E_B^- = \{(v, \ell) : \ell \in S_v^-\}$. For any $v \in V$, we also set $E_B^+(v) = \{\ell : (v, \ell) \in E_B^+\}$ and $E_B^-(v) = \{\ell : (v, \ell) \in E_B^-\}$.

The aforementioned framework can model various real-life situations and may be used to study the dynamics and the behavior of various network-like structures. For example, the individuals of a population may be viewed to correspond to vertices and that there is a (finite) *universal set of features* $\mathcal{M}$ corresponding to the set of ideals, beliefs, preferences etc. that an individual can perceive. For each feature $\ell \in \mathcal{M}$, we assume that every individual (vertex) can have either a positive, negative or indifferent *view/opinion* towards ℓ; in a way the set of views towards all the features of $\mathcal{M}$ can be seen as part of someone's personality. Relations between pairs of individuals are formed by taking into account their relative views over all features. In particular, if (say) an individual "likes" physics while another one does not, then there is a (small) *conflict* with respect to physics between them; otherwise there is an *agreement*. If agreements outnumber the conflicts, then there is a positive relation (i.e. positive edge) between those two individuals. Similarly, there is a negative edge between them if the conflicts outnumber the agreements, and there is no edge if conflicts are equal to the agreements.

In this paper, we initiate the study of random signed intersection graphs by providing several preliminary results concerning the number of conflicts among all vertices. As our main contribution, we study balance in $\mathcal{G}_{n,m,p,q}$ in the limiting case $n \to \infty$. In particular, in Theorem 4.3 we prove that, when $p = q = \Theta(1)$, and $m \geq 3$, $G_{n,m,p,q}$ is with high probability not balanced, since it contains $\Omega(n^3)$ negative 3-cycles (which is the maximum possible up to constant factors). Furthermore, we prove in Theorem 4.4 that Theorem 4.3 is tight in the following sense: if $m = \omega(\log n)$ and $|p-q|$ is bounded below by an arbitrarily small constant, $G_{n,m,p,q}$ is balanced with high probability. We also note that, if $m \leq 2$, $G_{n,m,p,q}$ is balanced with probability 1 (see a short explanation of this fact in the beginning of Section 4.1). On the other hand, if m is any constant larger than 2, and p, q are positive constants, the probability that $G_{n,m,p,q}$ has a negative cycle (and thus it is not balanced) is bounded by a constant. We finally note that, the difficulty in analyzing our model lies in the fact that edges are not independent (which is also the case in unsigned random intersection graphs). For our proofs, we exploit the special structure of the associated signed bipartite graph and we use coupling arguments and stochastic domination, as well as concentration bounds and several properties of the binomial distribution.

2 Useful Facts

In this section we mention some results that we use throughout the paper.

Theorem 2.1 (Multiplicative Chernoff Bound [12])**.** *Let $X_1, X_2, \cdots, X_n$ be independent Bernoulli random variables taking values in $\{0,1\}$. Let $X = \sum_{i=1}^{n} X_i$ and let $\mu = \mathbb{E}[X]$. Then the following hold for any $0 < \delta < 1$:*

(a) *Lower tail Chernoff bound*

$$\Pr(X \leq (1-\delta)\mu) \leq e^{-\frac{\delta^2\mu}{2}}.$$

(b) *Upper tail Chernoff bound*

$$\Pr(X \geq (1+\delta)\mu) \leq e^{-\frac{\delta^2\mu}{3}}.$$

We denote by $\mathcal{B}(n,p)$ the binomial distribution with parameters n, p. In particular, the above implies that $\Pr(|X - np| \leq \delta np) \geq 1 - 2e^{-\frac{\delta^2 np}{3}}$, $0 < \delta < 1$, for $X \sim \mathcal{B}(n,p)$.

The following lemma will be used for the proof of Lemma 2.2 below. Its obvious proof is omitted and the reader is referred to introductory books on probability (see e.g. [12]) and the exercises therein.

Lemma 2.1. *Let $X \sim \mathcal{B}\left(n, \frac{1}{2}\right)$ and $Y \sim \mathcal{B}\left(n', \frac{1}{2}\right)$ be binomial random variables with $n \geq n'$. Then, for any $0 \leq i' \leq i \leq \frac{n'}{2}$, $\frac{\Pr\left(X = \left\lceil \frac{n}{2} + i \right\rceil\right)}{\Pr\left(Y = \left\lfloor \frac{n'}{2} + i' \right\rfloor\right)} = \frac{\Pr\left(X = \left\lceil \frac{n}{2} - i \right\rceil\right)}{\Pr\left(Y = \left\lfloor \frac{n'}{2} - i' \right\rfloor\right)} \leq 1.$*

We could not find Lemma 2.2 in any textbook, and so we give the proof here.

Lemma 2.2. *Let* $X \sim \mathcal{B}\left(n, \frac{1}{2}\right)$ *and* $Y \sim \mathcal{B}\left(n', \frac{1}{2}\right)$ *be stochastically independent binomial random variables with* $n \geq n'$. *Then* $\Pr\left(\left|X - \frac{n}{2}\right| \geq \left|Y - \frac{n'}{2}\right|\right) > \frac{1}{2}$.

Proof: Notice that

$$\Pr\left(\left|X - \frac{n}{2}\right| \geq \left|Y - \frac{n'}{2}\right|\right) = \sum_{-\frac{n}{2} \leq a \leq \frac{n}{2}} \Pr\left(|a| \geq \left|Y - \frac{n'}{2}\right|\right) \Pr\left(X = \frac{n}{2} + a\right)$$

and also that $\Pr\left(|a| \geq \left|Y - \frac{n'}{2}\right|\right)$ is increasing with $|a|$. Let $Y' \sim \mathcal{B}\left(n', \frac{1}{2}\right)$ be stochastically independent of Y. Suppose that in the above sum we replace $\Pr\left(X = \frac{n}{2} + a\right)$ with $\Pr\left(Y' = \left\lfloor \frac{n'}{2} + a \right\rfloor\right)$ if $a \geq 0$ and with $\Pr\left(Y' = \left\lceil \frac{n'}{2} + a \right\rceil\right)$ if $a < 0$. Then by Lemma 2.1, we eliminate (i.e. set equal to 0) the multiplicative factors of probabilities $\Pr\left(|a| \geq \left|Y - \frac{n'}{2}\right|\right)$ with either $a < -\frac{n'}{2}$ or $a > \frac{n'}{2}$ (which equal to 1), while increasing the multiplicative factors of (smaller) probabilities $\Pr\left(|a| \geq \left|Y - \frac{n'}{2}\right|\right)$ with $-\frac{n'}{2} \leq a \leq \frac{n'}{2}$. Therefore,

$$\begin{aligned}
&\sum_{-\frac{n}{2} \leq a \leq \frac{n}{2}} \Pr\left(|a| \geq \left|Y - \frac{n'}{2}\right|\right) \Pr\left(X = \frac{n}{2} + a\right) \\
&\quad \geq \sum_{-\frac{n}{2} \leq a < 0} \Pr\left(|a| \geq \left|Y - \frac{n'}{2}\right|\right) \Pr\left(Y' = \left\lceil \frac{n'}{2} + a \right\rceil\right) \\
&\qquad + \sum_{0 \leq a \leq \frac{n}{2}} \Pr\left(|a| \geq \left|Y - \frac{n'}{2}\right|\right) \Pr\left(Y' = \left\lfloor \frac{n'}{2} + a \right\rfloor\right) \\
&\quad = \sum_{y'=0}^{n'} \Pr\left(\left|y' - \frac{n'}{2}\right| \geq \left|Y - \frac{n'}{2}\right|\right) \Pr(Y' = y) \\
&\quad = \Pr\left(\left|Y' - \frac{n'}{2}\right| \geq \left|Y - \frac{n'}{2}\right|\right). \qquad (1)
\end{aligned}$$

By symmetry, the above probability is at least $\frac{1}{2} + \frac{\Pr(Y=Y')}{2}$, completing the proof. □

We claim that a more general version of the lemma is true, where $\mathcal{B}\left(n, \frac{1}{2}\right)$ and $\mathcal{B}\left(n', \frac{1}{2}\right)$ are replaced with $\mathcal{B}(n, p)$ and $\mathcal{B}(n', p)$ respectively (notice that the probability of success is the same for both binomial distributions). However, this version of the lemma suffices for our purposes in this paper.

3 Preliminary results

The following result states that, by definition, the distribution $\mathcal{G}_{n,m,p,q}$ is symmetric with respect to p and q. In particular:

Lemma 3.1. *For any positive p, q, with $p + q \leq 1$, the distributions of $G_{n,m,p,q}$ and $G_{n,m,q,p}$ are identical.*

Proof: We couple the two models in such a way that whenever there is a positive (respectively, negative) feature choice in $G_{n,m,p,q}$, there is a negative (respectively, positive) feature choice in $G_{n,m,q,q}$. Then, by definition, the graph instance $G_{n,m,p,q}$ is identical to $G_{n,m,q,p}$ and have the same probability. □

The above result will be useful in subsection 3.1 when proving various domination results for the case $q = 1 - p$. In particular, because of Lemma 3.1, we only need to consider values for p that are upper bounded by $\frac{1}{2}$.

3.1 Negativity

We define the *negativity* of an edge (u, v) as the inner product $N_{uv} \stackrel{def}{=} \sum_\ell \mathbf{x}_v[\ell]\mathbf{x}_u[\ell]$. Similarly, we define the negativity of $G_{n,m,p,q}$ as follows:

$$N_G \stackrel{def}{=} \sum_{u \neq v} N_{uv}. \tag{2}$$

We prove an intuitive stochastic domination[1] result relating p and N_G (Corollary 3.1). Loosely speaking, the closer p is to 1/2, the smaller N_G will be (statistically speaking). For the proof, we need the following definition, which is a generalization of the random signed intersection graphs model. In particular, we define the model $\mathcal{G}_{n,m,\mathbf{P},\mathbf{Q}}$ as follows: Let $\mathcal{V}, \mathcal{M}$ be the sets of vertices and features respectively. Let also $\mathbf{P}$ (respectively, $\mathbf{Q}$) be a $n \times m$ matrix, where $\mathbf{P}_{v,\ell} \in [0, 1]$ (respectively, $\mathbf{Q}_{v,\ell} \in [0, 1 - \mathbf{P}_{v,\ell}]$) is the probability that $\mathbf{x}_v[\ell] = +1$ (respectively, $\mathbf{x}_v[\ell] = -1$). As usual, two vertices u, v of a random instance of $\mathcal{G}_{n,m,\mathbf{P},\mathbf{Q}}$ are connected with a *positive edge* if and only if $\sum_\ell \mathbf{x}_v[\ell]\mathbf{x}_u[\ell] > 0$, with a *negative edge* if $\sum_\ell \mathbf{x}_v[\ell]\mathbf{x}_u[\ell] < 0$ and not connected if $\sum_\ell \mathbf{x}_v[\ell]\mathbf{x}_u[\ell] = 0$.

We first prove the following result on the negativity of a random instance $G_{n,m,\mathbf{P},\mathbf{1}_{n\times m}-\mathbf{P}}$, where $\mathbf{1}_{n\times m}$ denotes the all-ones $n \times m$ matrix:

Lemma 3.2. *Let $0 < p < p' \leq \frac{1}{2}$ and let $\mathbf{P}$ be such that $\mathbf{P}_{v,\ell} \in \{p, p'\}$ (i.e. takes only two possible values). Let also $\mathbf{P}'$ be equal to $\mathbf{P}$, except for one pair of indices $(v_0, \ell_0) \in \mathcal{V} \times \mathcal{M}$, where $\mathbf{P}_{v_0,\ell_0} = p$ and $\mathbf{P}'_{v_0,\ell_0} = p'$. If $G \sim \mathcal{G}_{n,m,\mathbf{P},\mathbf{1}_{n\times m}-\mathbf{P}}$ and $G' \sim \mathcal{G}_{n,m,\mathbf{P}',\mathbf{1}_{n\times m}-\mathbf{P}'}$, then N_G stochastically dominates $N_{G'}$, i.e. $N_G \geq_{st} N_{G'}$.*

Proof: By definition, we have that

[1]For two random variables X, Y taking values in the same set $\mathcal{A} \subseteq \mathbf{R}$, we say that X stochastically dominates Y and we write $X \geq_{st} Y$ if and only if $\Pr(X \geq k) \geq \Pr(Y \geq k)$, for any $k \in \mathcal{A}$.

$$\begin{aligned} N_G &= \sum_{u \neq v} \sum_{\ell} \mathbf{x}_v[\ell]\mathbf{x}_u[\ell] = \sum_{(v,\ell)} \mathbf{x}_v[\ell] \sum_{u \neq v} \mathbf{x}_u[\ell] \\ &= \mathbf{x}_{v_0}[\ell_0] \sum_{u \neq v_0} \mathbf{x}_u[\ell] + \sum_{(v,\ell)\neq(v_0,\ell_0)} \mathbf{x}_v[\ell] \sum_{u \neq v} \mathbf{x}_u[\ell]. \end{aligned} \tag{3}$$

Therefore, by coupling $\mathcal{G}_{n,m,\mathbf{P},\mathbf{1}_{n\times m}-\mathbf{P}}$ and $\mathcal{G}_{n,m,\mathbf{P'},\mathbf{1}_{n\times m}-\mathbf{P'}}$, we only need to prove that $X \stackrel{def}{=} \mathbf{x}_{v_0}[\ell_0] \sum_{u \neq v_0} \mathbf{x}_u[\ell]$ stochastically dominates $X' \stackrel{def}{=} \mathbf{x}'_{v_0}[\ell_0] \sum_{u \neq v_0} \mathbf{x}_u[\ell]$, where $\Pr(\mathbf{x}_{v_0}[\ell_0] = +1) = p$ and $\Pr(\mathbf{x}'_{v_0}[\ell_0] = +1) = p'$.

Notice now that, for any $-n+1 \leq k \leq n-1$, by independence of feature choices,

$$\begin{aligned} \Pr(X \geq k) &= \Pr(\mathbf{x}_{v_0}[\ell_0] = +1) \Pr\left(\sum_{u \neq v_0} \mathbf{x}_u[\ell] \geq k\right) \\ &\quad + \Pr(\mathbf{x}_{v_0}[\ell_0] = -1) \Pr\left(\sum_{u \neq v_0} \mathbf{x}_u[\ell] \leq -k\right) \\ &= p\left(\Pr\left(\sum_{u \neq v_0} \mathbf{x}_u[\ell] \geq k\right) - \Pr\left(\sum_{u \neq v_0} \mathbf{x}_u[\ell] \leq -k\right)\right) \\ &\quad + \Pr\left(\sum_{u \neq v_0} \mathbf{x}_u[\ell] \leq -k\right). \end{aligned}$$

Since $p < p' \leq \frac{1}{2}$, we have that $\Pr\left(\sum_{u \neq v_0} \mathbf{x}_u[\ell] \geq k\right) < \Pr\left(\sum_{u \neq v_0} \mathbf{x}_u[\ell] \leq -k\right)$, and so $\Pr(X \geq k)$ is a decreasing function of p. In particular, $\Pr(X \geq k) > \Pr(Y \geq k)$, for any $-n+1 \leq k \leq n-1$, which completes the proof. □

Applying the above lemma inductively and noting that $\mathcal{G}_{n,m,p,1-p}$ is identical to $\mathcal{G}_{n,m,1-p,p}$ and also to $\mathcal{G}_{n,m,\mathbf{P},\mathbf{1}_{n\times m}-\mathbf{P}}$, with $\mathbf{P}_{v,\ell} = p$, for each $v \in \mathcal{V}, \ell \in \mathcal{M}$, we get the following:

Corollary 3.1. *Let $G \sim \mathcal{G}_{n,m,p,1-p}$ and $G' \sim \mathcal{G}_{n,m,p',1-p'}$ be two random instances of the random signed intersection graphs model, where $|2p-1| \geq |2p'-1|$. Then $N_G \geq_{st} N_{G'}$. In particular, the negativity of a random instance of $\mathcal{G}_{n,m,p,1-p}$ is statistically minimal when $p = \frac{1}{2}$.*

We conjecture that an analog to Corollary 3.1 also holds for the number of negative edges:

Conjecture 3.1. *Let $G_{n,m,p,1-p}$ and $G_{n,m,p',1-p'}$ be two random instances of the random signed intersection graphs model, where $|2p-1| \geq |2p'-1|$. Let also X and Y denote the number of negative edges in $G_{n,m,p,1-p}$ and $G_{n,m,p',1-p'}$ respectively. Then $X \leq_{st} Y$, i.e. $\Pr(X \geq k) \leq \Pr(Y \geq k)$, for any k.*

It is easy to see that the above conjecture holds in the simple case of 2 vertices, say u, v. Indeed, setting $a_p \stackrel{def}{=} 2p(1-p)$, the probability that (u, v) is negative is

$$\Pr(\sigma_G(uv) = -1) = \sum_{\ell > m/2} \binom{m}{\ell} a_p^{\ell}(1-a_p)^{m-\ell}.$$

Since the above probability is an increasing function of a_p and a_p is a convex function of p within $(0, 1)$, the maximum of $\Pr(\sigma_G(uv) = -1)$ is attained for $p = 1/2$, for which a_p is maximized.

By a similar proof one can also show that Conjecture 3.1 holds for the simple case $m = 2$.

4 Balance in $\mathcal{G}_{n,m,p,q}$

In a signed graph, a *negative cycle* is a cycle that contains an odd number of negative edges. If the signed graph does not contain any negative cycle, we say that it is *balanced*.

We first note that, when $p + q$ is very small, then $G_{n,m,p,q}$ is balanced, since (forgetting edge signs) it has not cycles whatsoever.

Theorem 4.1. *If $p + q = o\left(\frac{1}{\min\{n,m\}}\right)$, $G_{n,m,p,q}$ is balanced with high probability.*

Proof: Notice that the existence of a negative cycle of size larger than $\min\{n, m\}$ in $G_{n,m,p,q}$ implies the existence of a cycle of size at most $2\min\{n, m\}$ in $B_{n,m,p,q}$. But, by the union bound, and because of independence of feature choices, the probability that $B_{n,m,p,q}$ has a cycle (either negative or positive) of size $2k$ is upper bounded by $n^k m^k (p+q)^{2k} \stackrel{def}{=} x^k$.

Therefore, the probability that $G_{n,m,p,q}$ has a negative cycle is at most

$$\sum_{k=2}^{\min\{n,m\}} x^k = \frac{x^{\min\{n,m\}} - x^2}{x - 1}.$$

For $p + q = o\left(\frac{1}{\min\{n,m\}}\right)$, we have $x = o(1)$, which completes the proof. □

In view of the above, in what follows we consider balance of $\mathcal{G}_{n,m,p,q}$ in the more interesting case $p + q = \Theta(1)$. We leave the case $o\left(\frac{1}{\min\{n,m\}}\right) = p + q = o(1)$ as an open problem.

4.1 The case $p + q = \Theta(1)$

In this section we prove our main results for negative cycles of size 3 in $G_{n,m,p,q}$ when $p + q = \Theta(1)$, i.e. $p + q$ is bounded below by a constant. We focus on the case $m \geq 3$, since any $G_{n,m,p,q}$ is balanced if $m \leq 2$. Indeed, when $m \leq 2$ we can define the bi-partition $A = \{v : \mathbf{x}_v[\ell] = +1, \forall \ell \in \mathcal{M}\}$ and $B = V \backslash A$; observe now that all negative edges have one endpoint in A and one in B and all positive edges either have both points in A or both in B, which implies that all cycles have an even number of negative edges.

We first present our proof for the special case where $p + q = 1$ and then show how this can be used for the general case $p + q = \Theta(1)$. In particular, we prove the following:

Theorem 4.2. *Let $p = q = \frac{1}{2}$ and $m \geq 3$. Then $G_{n,m,p,q}$ has $\Omega(n^3)$ negative cycles with high probability as n goes to infinity.*

Proof: Fix distinct vertices $x, y, z \in V$. We will bound the probability that $\sigma(yz) = -1$ and $\sigma(xy) = \sigma(xz) = +1$. Notice that, by symmetry, this probability is equal to the probability that all edges are negative; indeed, alternating the signs of the feature choices of x results in switching x (i.e. changing the signs of all incident edges of x).

Define $K \stackrel{def}{=} |E_B^+(z) \cap E_B^+(y)| + |E_B^-(z) \cap E_B^-(y)|$ to be the random variable that counts the number of features where z and y agree. Clearly, $K \sim \mathcal{B}(m, p^2 + q^2) = \mathcal{B}\left(K, \frac{1}{2}\right)$.

Define also $X \stackrel{def}{=} |E_B^+(z) \cap E_B^+(y) \cap E_B^+(x)| + |E_B^-(z) \cap E_B^-(y) \cap E_B^-(x)|$ to be the random variable that counts the number of features where z, y and x agree. Clearly, $X_{|K} \sim \mathcal{B}\left(K, \frac{1}{2}\right)$, i.e., the distribution of X given K is binomial with parameters K and $\frac{1}{2}$.

Finally, define $Y \stackrel{def}{=} |E_B^-(z) \cap E_B^+(y) \cap E_B^+(x)| + |E_B^+(z) \cap E_B^-(y) \cap E_B^-(x)|$ to be the random variable that counts the number of features where only y and x agree (and not z). Clearly, $Y_{|K} \sim \mathcal{B}\left(m - K, \frac{1}{2}\right)$, i.e., the distribution of Y given K is binomial with parameters $m - K$ and $\frac{1}{2}$. Furthermore, given K the random variables $X_{|K}$ and $Y_{|K}$ are stochastically independent.

By the above definitions, first note that

$$\Pr(\sigma(yz) = -1) = \Pr\left(K < \frac{m}{2}\right). \tag{4}$$

Furthermore, to have $\sigma(xz) = +1$ and $\sigma(xy) = +1$ at the same time, we need $X - (K - X) + \min\{Y, m - K - Y\} - (m - K - \min\{Y, m - K - Y\}) > 0$, or equivalently $m - 2X < 2Y < m - 2(K - X)$. Indeed (please refer to Figure 1), to have $\sigma(x, z) = +1$, we must have $0 < \sum_\ell \mathbf{x}_x[\ell]\mathbf{x}_z[\ell] = X - (K - X) - Y + m - K - Y = m - 2(K - X) - 2Y$. Furthermore, to have $\sigma(xy) = +1$, we must have $0 < \sum_\ell \mathbf{x}_x[\ell]\mathbf{x}_y[\ell] = X - (K - X) + Y - (m - K - Y) = 2Y - (m - 2X)$. Therefore,

$$\begin{aligned}
\Pr(\sigma(xz) = +1, \sigma(xy) = +1|K) &= \Pr(m - 2X < 2Y < m - 2(K - X)|K) \\
&= \Pr\left(\frac{m}{2} - X < Y < \frac{m}{2} - K + X \middle| K\right) \\
&= \Pr\left(|Y - \mathbb{E}[Y|K]| < X - \mathbb{E}[X|K]|K\right) \qquad (5)
\end{aligned}$$

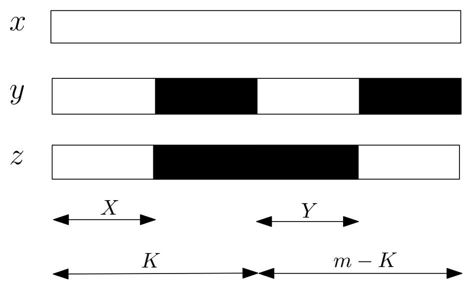

Figure 1: An instance of the feature choices of 3 vertices x, y, z. Same color means that the corresponding feature choices are the same. For simplicity, feature choices have been grouped.

where in the last equation we used $\mathbb{E}[X|K] = \frac{K}{2}$ and $\mathbb{E}[Y|K] = \frac{m-K}{2}$. Clearly, the above probability is non-zero only if $2X > K$.

We now prove the following bound on the probability that $\sigma(yz) = -1$, $\sigma(xz) = +1$ and $\sigma(xy) = +1$. We make no attempt to optimize the various constants coming from concentration bounds.

Claim. *For $p = q = \frac{1}{2}$, and any $m \geq 3$, there is a constant $\gamma \in (0, 1)$, such that*

$$\Pr\left(K < \frac{m}{2}, |Y - \mathbb{E}[Y]| < X - \mathbb{E}[X]\right) \geq \gamma, \quad \text{for any } m \geq 3. \tag{6}$$

Proof: We will assume without loss of generality that m is sufficiently large (e.g. $m \geq 1000$), since the claim is obvious for small values of m.

We first note that with at least constant probability, we have $K = \Theta(m)$. In particular, since $X \sim \mathcal{B}\left(m, \frac{1}{2}\right)$, by symmetry,

$$\begin{aligned}\Pr\left(\frac{4m}{10} \leq K < \frac{m}{2}\right) &\geq \frac{1}{2}\Pr\left(\left|K - \frac{m}{2}\right| \leq \frac{m}{10}\right) - \Pr\left(K = \left\lceil\frac{m}{2}\right\rceil\right) \\ &\geq \frac{1}{2} - \exp\left\{-\frac{m}{400}\right\} + \binom{m}{\frac{m}{2}}2^{-m} \geq \frac{1}{3}.\end{aligned} \tag{7}$$

where in the second inequality we applied the Chernoff bound, and the last inequality holds for all large enough m.

Second, we prove that, given $K = \Theta(m)$, with at least constant probability we have $|Y - \mathbb{E}[Y]| < X - \mathbb{E}[X]$. We first note that, by symmetry of $\mathcal{B}\left(K, \frac{1}{2}\right)$ and $\mathcal{B}\left(m - K, \frac{1}{2}\right)$ around their expected values, we have that

$$\Pr\left(\left|Y - \frac{m-K}{2}\right| < X - \frac{K}{2}\middle|K\right) = \frac{1}{2}\Pr\left(\left|Y - \frac{m-K}{2}\right| < \left|X - \frac{K}{2}\right|\middle|K\right). \tag{8}$$

Notice also that we can write $Y \stackrel{def}{=} X' + Z$, where $X' \sim \mathcal{B}\left(K, \frac{1}{2}\right)$ is stochastically independent of X and $Z \sim \mathcal{B}\left(m - 2K, \frac{1}{2}\right)$ is stochastically independent of both X and X'.

Finally, note that, among all the feature choices that satisfy $\left|Y - \frac{m-K}{2}\right| < \left|X - \frac{K}{2}\right|$, for at least half of them we also have $\left|Y - \frac{m-K}{2}\right| = \left|\left|X' - \frac{K}{2}\right| - \left|Z - \frac{m-2K}{2}\right|\right|$; indeed, flipping all feature choices corresponding to Z produces a distinct, equiprobable set of feature choices. Therefore, by (8), we get

$$\Pr\left(\left|Y - \frac{m-K}{2}\right| < X - \frac{K}{2}\middle|K\right) \\ \geq \frac{1}{4}\Pr\left(\left|\left|X' - \frac{K}{2}\right| - \left|Z - \frac{m-2K}{2}\right|\right| < \left|X - \frac{K}{2}\right|\middle|K\right) \tag{9}$$

Notice now that, for $\frac{4m}{10} \leq K < \frac{m}{2}$ (which is the event from inequality (7)), we have that $K > m - 2K$. Therefore, by Lemma 2.2, with probability at least $\frac{1}{2}$ we have $\left|Z - \frac{m-2K}{2}\right| \leq \left|X' - \frac{K}{2}\right|$. Furthermore, by symmetry, given this event (call it E), it is easy to see that with probability at least $\frac{1+\Pr(X'=X|E)}{2} > \frac{1}{2}$, we will have $\left|X' - \frac{K}{2}\right| \leq \left|X - \frac{K}{2}\right|$. Combining the above with (9), we then have

$$\Pr\left(\left|Y - \frac{m-K}{2}\right| < X - \frac{K}{2}\middle|\frac{4m}{10} \leq K < \frac{m}{2}\right) \geq \frac{1}{16}. \tag{10}$$

Finally, by the chain rule and equations (7) and (10), we get

$$\Pr\left(K < \frac{m}{2}, |Y - \mathbb{E}[Y]| < X - \mathbb{E}[X]\right) \geq \frac{1}{48},$$

which completes the proof of the claim. □

By the above claim, $\Pr(xyz \text{ induce a negative cycle}) \geq \gamma$. Now fix an ordering $H_1, H_2, \ldots, H_{\binom{n}{3}}$ of all subsets of vertices of size 3 and, for $1 \leq i \leq \binom{n}{3}$, define T_i to be the indicator variable that the i-th triplet of vertices (i.e. H_i) induces a negative cycle. In particular, we have that $\Pr(T_i = 1) \geq \gamma$, for each i. Let also T denote the number of negative cycles in $G_{n,m,p,1-p}$. Then, by linearity of expectation, we have that $\mathbb{E}[T] = \sum_{i=1}^{\binom{n}{3}} \gamma \geq \binom{n}{3}\gamma = \Theta(n^3)$, which goes to infinity with n.

It remains to show that T is also concentrated around its expected value. To prove this, we employ a technique described in [1]. For indices i, j, we write $i \sim j$ if the events $\{T_i = 1\}$ and $\{T_j = 1\}$ are not independent. Notice now that the events $\{T_i = 1\}, 1 \leq i \leq \binom{n}{3}$ are symmetric, i.e. for any $i \neq j$, there is an automorphism of the underlying probability space that sends event $\{T_i = 1\}$ to $\{T_j = 1\}$. Now fix i, and define $\Delta^* = \sum_{j \sim i} \Pr(T_j = 1 | T_i = 1)$. We will prove that $\Delta^* = o(\mathbb{E}[T])$, from which the desired concentration follows from the following result from [1]:

Corollary 4.1 ([1], chapter 4)**.** *If $\mathbb{E}[T] \to \infty$ and $\Delta^* = o(\mathbb{E}[T])$ then $T > 0$ almost always. Furthermore, $T \sim \mathbb{E}[T]$ almost always.*

Notice that, for any i, j, the events $\{T_i = 1\}$ and $\{T_j = 1\}$ are stochastically independent when the corresponding triplets of vertices H_i and H_j are disjoint. However, we can prove independence also in the case $|H_i \cap H_j| = 1$. Towards this end, suppose that $H_i \cap H_j = v$ (i.e. v is the common vertex of H_i and H_j) and let $\mathbf{x}_v$ be the vector of feature choices for v; since $p = q = \frac{1}{2}$, all values in $\mathbf{x}_v$ are non-zero. The crucial observation is that, $\Pr\{T_i = 1|\mathbf{x}_v\} = \Pr\{T_i = 1|\mathbf{x}_v[\ell] = 1, \forall \ell \in \mathcal{M}\}$, i.e. the event $\{T_i = 1\}$ is stochastically independent of the specific feature choices of v. Indeed, let $H_i = \{v, u, w\}$ and let $\mathbf{x}_u, \mathbf{x}_w$ be the feature choices of the other two vertices in H_i respectively. Also let F (respectively, F') be the set of vectors $\mathbf{x}_u, \mathbf{x}_w$ such that H_i induces a negative cycle given $\mathbf{x}_v$ (respectively, given $\mathbf{x}_v[\ell] = 1, \forall \ell \in \mathcal{M}$). Then we can define a one-to-one correspondence $f : F \to F'$ between F and F' by $f(\mathbf{x}_u, \mathbf{x}_w) = (\mathbf{x}_u \odot \mathbf{x}_v, \mathbf{x}_w \odot \mathbf{x}_v)$, where $\odot$ denotes component-wise multiplication. A similar argument shows that the event $\{T_j = 1\}$ is stochastically independent of the specific feature choices of v.

In view of the above, and also taking into account that, $i \nsim j$ if $|H_i \cap H_j| = 3$, we conclude that $i \sim j$ only if $|H_i \cap H_j| = 2$. Therefore, for fixed i, there are $3(n-3)$ indices j such that $j \sim i$, which gives $\Delta^* = O(n) = o(\mathbb{E}[T])$, and the proof is complete. □

We now consider the more general case $1 \leq p + q = \Theta(1)$.

Theorem 4.3. *Let $\frac{1}{2} \geq p = q \geq c$, for some positive constant $c \leq \frac{1}{2}$, and let $m \geq 3$. Then $G_{n,m,p,q}$ has $\Theta(n^3)$ negative cycles with high probability.*

Proof: Fix distinct vertices $x, y, z \in V$. We will use the Claim in the proof of Theorem 4.2 to prove a constant lower bound on the probability that $\{x, y, z\}$ induce a negative cycle in $G_{n,m,p,q}$. Towards this end, let $\mathcal{M}_{xyz}$ denote the set of features that have been selected (either with $+1$ or -1) by all three vertices. Furthermore, let $\mathcal{M}_{xy\bar{z}}$ denote the set of features that have been selected by x and y but not z. Let $\mathcal{M}_{\bar{x}yz}$ and $\mathcal{M}_{x\bar{y}z}$ be defined similarly. Note that $|\mathcal{M}_{xyz}| \sim \mathcal{B}(m, (p+q)^3)$ and each of $|\mathcal{M}_{xy\bar{z}}|, |\mathcal{M}_{\bar{x}yz}|, |\mathcal{M}_{x\bar{y}z}|$ follows $\mathcal{B}(m, (p+q)^2(1-p-q))$; note that these random variables are not independent.

Consider now the random signed intersection graph instance G_0 with associated signed bipartite graph B_0 having vertex set $\{x, y, z\} \cup \mathcal{M}_{x,y,z}$ and edge set $E^+_{B_0} \cup E^-_{B_0}$, where $E^+_{B_0} = \{(v, \ell) : \ell \in S^+_v \cap \mathcal{M}_{x,y,z}, v \in \{x, y, z\}\}$ and $E^-_{B_0} = \{(v, \ell) : \ell \in S^-_v \cap \mathcal{M}_{x,y,z}, v \in \{x, y, z\}\}$. In particular, G_0 is the signed intersection graph instance constructed by considering only vertices x, y, z and the set of features that have been selected by all three vertices. Notice then that, given $\mathcal{M}_{x,y,z}$, G_1 is distributed as $G_{3,|\mathcal{M}_{x,y,z}|,\frac{1}{2},\frac{1}{2}}$. Furthermore, since $\Pr(|\mathcal{M}_{x,y,z}| \geq 3) \geq c^3$, we can apply the Claim from the proof of Theorem 4.2 to conclude that, with (constant) probability at least $c^3\gamma$ (where γ is the constant defined in the Claim), G_0 will be a negative cycle.

We now show that, with constant probability the remaining set of features does not reduce this probability by more than a multiplicative constant factor. To prove this, we define the random signed intersection graph instances G_1, G_2, G_3 similarly to G_0; in particular G_1 (respectively, G_2, G_3) is the signed intersection graph instance constructed by considering only vertices x, y, z and the set of features $\mathcal{M}_{xy\bar{z}}$ (respectively, $\mathcal{M}_{\bar{x}yz}, \mathcal{M}_{x\bar{y}z}$). Notice that, given $\mathcal{M}_{xy\bar{z}}$, G_1 induces at most a single edge between x, y, which is distributed as $G_{2,|\mathcal{M}_{xy\bar{z}}|,\frac{1}{2},\frac{1}{2}}$. Similarly, given $\mathcal{M}_{\bar{x}yz}$, G_2 has at most one edge between y, z, which is distributed as $G_{2,|\mathcal{M}_{\bar{x}yz}|,\frac{1}{2},\frac{1}{2}}$, and, given $\mathcal{M}_{x\bar{y}z}$, G_3 has at most one edge between x, z, which is distributed as $G_{2,|\mathcal{M}_{x\bar{y}z}|,\frac{1}{2},\frac{1}{2}}$. The crucial observation is that, given the set of features $\mathcal{M}_{x,y,z}, \mathcal{M}_{xy\bar{z}}, \mathcal{M}_{\bar{x}yz}, \mathcal{M}_{x\bar{y}z}$ the signs $\sigma_1(xy), \sigma_2(yz), \sigma_3(xz) \in \{0, \pm 1\}$ of edges in G_1, G_2, G_3 are mutually independent and are independent of the signs of edges $\sigma_0(xy), \sigma_0(yz), \sigma_0(xz)$ in G_0. Furthermore, by independence and by the definition of G_1, we have

$$\Pr(\sigma_1(xy)\sigma_0(xy) \geq 0) \geq \frac{1}{2}.$$

In particular, this inequality implies that, with constant probability the sign $\sigma(xy)$ of the edge between x, y in the original instance $G_{n,m,p,q}$ will be equal to $\sigma_0(x, y)$ (i.e. the sign in G_0). Identical inequalities hold for $\Pr(\sigma_2(yz)\sigma_0(yz) \geq 0)$ and $\Pr(\sigma_3(xz)\sigma_0(xz) \geq 0)$. Putting this all together, if $\mathcal{E}$ is the event $\{xyz$ induce a negative cycle in $G_{n,m,p,q}\}$ we have

$$\begin{aligned}
\Pr(\mathcal{E}) &\geq c^3 \Pr(\mathcal{E} | \{|\mathcal{M}_{x,y,z}| \geq 3\}) \\
&\geq c^3 \frac{1}{2^3} \Pr(\mathcal{E} | \{\sigma_1(xy)\sigma_0(xy) \geq 0, \sigma_2(yz)\sigma_0(yz) \geq 0, \sigma_3(xz)\sigma_0(xz) \geq 0\}, \\
&\qquad \{|\mathcal{M}_{x,y,z}| \geq 3\}) \qquad (11) \\
&= c^3 \frac{1}{2^3} \Pr(xyz \text{ induce a negative cycle in } G_0 | \{|\mathcal{M}_{x,y,z}| \geq 3\}) \qquad (12) \\
&\geq c^3 \frac{1}{2^3} \gamma,
\end{aligned}$$

where γ is the constant from the Claim in the proof of Theorem 4.2. In particular, inequality (11) above follows by mutual independence of $\sigma_1(xy)$, $\sigma_2(yz)$, $\sigma_3(xz)$, $\sigma_0(xy)$, $\sigma_0(yz)$, $\sigma_0(xz)$; notice also that $\{|\mathcal{M}_{x,y,z}|$ can only affects these signs indirectly by forcing some of the feature sets $\mathcal{M}_{xy\bar{z}}, \mathcal{M}_{\bar{x}yz}, \mathcal{M}_{x\bar{y}z}$ to be empty; however, in this case the corresponding signs are equal to 0, which works in our favor. Finally, equality (12) follows by observing that the when the signs of edges xy, yz, xz are determined by G_0, $\mathcal{E}$ is identical to the event $\{xyz$ induce a negative cycle in $G_0\}$.

We now proceed similarly to the proof of Theorem 4.2. In particular, let T denote the expected number of negative cycles in $G_{n,m,p,q}$. In view of the above, we have $\mathbb{E}[T] = \Theta(n^3)$. To prove concentration of T around its expected value, we apply Corollary 4.1; for simplicity we use the same notation. The main difference here is that, since $p + q$ can be smaller than 1, we can no longer claim that the events $\{T_i = 1\}$ and $\{T_j = 1\}$ are stochastically independent when $|H_i \cap H_i| = 1$. Nevertheless, for fixed i, there are $3\binom{n-3}{2}$ indices j such that $|H_i \cap H_i| = 1$, and so $\Delta^* = O(n^2) = o(\mathbb{E}[T])$ as needed. □

It is easy to see that if m is small and both p and q are constants, then the probability that $G_{n,m,p,q}$ has a 3-negative cycle is large enough (for example, if m is also constant, this probability is constant as well). On the other hand, when m is sufficiently large, we prove the following:

Theorem 4.4. *Let $m = \omega(\log n)$. If $|p - q| \geq \epsilon$, for any arbitrarily small positive constant $0 < \epsilon \leq 1$, then $G_{n,m,p,q}$ is balanced with high probability as $n \to \infty$.*

Proof: We assume without loss of generality that $p > q$. For any two vertices v, u, let $\mathcal{M}_{vu}$ denote the set of features that have been selected (either with $+1$ or -1) by both vertices; note that $|\mathcal{M}_{vu}| \sim \mathcal{B}(m, (p+q)^2)$. Therefore, by the lower tail Chernoff bound, setting $\delta = \frac{1}{2}$, we have

$$\Pr\left(|\mathcal{M}_{vu}| \leq \frac{1}{2}m(p+q)^2\right) \leq e^{-\frac{m(p+q)^2}{8}} \leq e^{-\frac{m\epsilon^2}{8}} = o(n^{-2}), \tag{13}$$

where the second inequality follows from the fact $p + q \geq \epsilon$ and the final equation follows from the fact $m = \omega(\log n)$. In particular, the above inequality states that $|\mathcal{M}_{vu}| > \frac{1}{2}m(p+q)^2 \geq \frac{m\epsilon^2}{2} = \omega(\log n)$ with probability $1 - o(n^{-2})$. Notice now that, by definition, given $\mathcal{M}_{vu}$, the sign of vu in $G_{n,m,p,q}$ is distributed as the sign of the (single) edge in a random signed intersection graph instance $G_{2,|\mathcal{M}_{vu}|,\frac{p}{p+q},\frac{q}{p+q}}$, i.e. with 2 vertices, $|\mathcal{M}_{vu}|$ features and $+$ (respectively, $-$) selection probability $\frac{p}{p+q}$ (respectively, $\frac{q}{p+q}$). Indeed, the latter follows by independence and by observing that $\Pr(\ell \in S_v^+ | \ell \in S_v^+ \cup S_v^-) = \frac{p}{p+q}$. Therefore, given $\mathcal{M}_{vu}$, the number C_{vu} of features on which v, u agree follows $\mathcal{B}\left(|\mathcal{M}_{vu}|, \frac{p^2}{(p+q)^2} + \frac{q^2}{(p+q)^2}\right)$. In particular, we have $\mathbb{E}\left[C_{vu} \big| |\mathcal{M}_{vu}|\right] = |\mathcal{M}_{vu}| \left(\frac{p^2}{(p+q)^2} + \frac{q^2}{(p+q)^2}\right) = |\mathcal{M}_{vu}| \left(\frac{1}{2} + \frac{1}{2}\left(\frac{p-q}{p+q}\right)^2\right) \geq |\mathcal{M}_{vu}| \left(\frac{1}{2} + \frac{1}{2}\epsilon^2\right)$. By the lower tail Chernoff bound, we then have, for any $0 < \delta < 1$,

$$\Pr\left(C_{vu} \leq (1-\delta)|\mathcal{M}_{vu}| \left(\frac{p^2}{(p+q)^2} + \frac{q^2}{(p+q)^2}\right) \Big| |\mathcal{M}_{vu}|\right) \leq e^{-\frac{\delta^2(1+\epsilon^2)|\mathcal{M}_{vu}|}{8}}. \tag{14}$$

Selecting $\delta = \frac{\epsilon^2}{1+\epsilon^2}$ we have that

$$(1-\delta)|\mathcal{M}_{vu}| \left(\frac{p^2}{(p+q)^2} + \frac{q^2}{(p+q)^2}\right) \geq (1-\delta)|\mathcal{M}_{vu}| \left(\frac{1}{2} + \frac{1}{2}\epsilon^2\right) = \frac{|\mathcal{M}_{vu}|}{2}.$$

Therefore, we have

$$\begin{aligned} \Pr(\sigma(vu) = -1 \big| |\mathcal{M}_{vu}|) &= \Pr\left(C_{vu} < \frac{|\mathcal{M}_{vu}|}{2} \Big| |\mathcal{M}_{vu}|\right) \\ &\leq \Pr\left(C_{vu} < (1-\delta)|\mathcal{M}_{vu}| \left(\frac{p^2}{(p+q)^2} + \frac{q^2}{(p+q)^2}\right) \Big| |\mathcal{M}_{vu}|\right) \\ &\leq e^{-\frac{\epsilon^2|\mathcal{M}_{vu}|}{8}}, \end{aligned} \tag{15}$$

where the last inequality follows from (14) by setting $\delta = \frac{\epsilon^2}{1+\epsilon^2}$. We can now bound the probability that vu is negative as follows:

$$\begin{aligned} \Pr(\sigma(vu) = -1) &\leq \Pr\left(|\mathcal{M}_{vu}| \leq \frac{1}{2}m(p+q)^2\right) \\ &\quad + \Pr\left(\sigma(vu) = -1 \Big| |\mathcal{M}_{vu}| > \frac{1}{2}m(p+q)^2\right) \qquad (16) \\ &\leq o(n^{-2}) + e^{-\frac{\epsilon^2 m(p+q)^2}{16}} = o(n^{-2}), \qquad (17) \end{aligned}$$

for all $m = \omega(\log n)$, where the second inequality follows by (13) and (15). By the union bound, we then have that, with probability that goes to 1 as $n \to \infty$, any pair of vertices agrees in more than $m/2$ features. This implies that all edges are positive, and so $G_{n,m,p,q}$ is balanced with high probability as needed. □

5 Further Research

In this paper we defined and initiated the study of the random signed intersection graphs model $\mathcal{G}_{n,m,p,q}$. As our main contribution, we proved that, if $p = q = \Theta(1)$, then with high probability $G_{n,m,p,q}$ is not balanced. On the other hand, if $m = \omega(\log n)$ and $|p-q|$ is lower bounded by some arbitrarily small constant, then $G_{n,m,p,q}$ is not balanced with high probability. A natural open problem related to balance of $G_{n,m,p,q}$ is to determine the *line index*, namely the smallest number of edges whose inversion of signs results in a balanced graph. By Theorems 4.3 and 4.4 we get the following:

Corollary 5.1. *Let p, q be such that $1 \geq p + q \geq c$, for some positive constant c. If $m \geq 3$ and $p = q$, then the line index of $G_{n,m,p,q}$ is $\Theta(n^2)$ with high probability. On the other hand, if $m = \omega(\log n)$ and $|p - q| \geq \epsilon$, for any arbitrarily small positive constant $0 < \epsilon \leq 1$, then the line index of $G_{n,m,p,q}$ is equal to 0 with high probability.*

We note that the proof of our result concerning the imbalance of random instances of the model $\mathcal{G}_{n,m,p,q}$ with $p = q$ relies on a rough lower bound of the probability that a fixed triplet of vertices induces a negative cycle (see the Claim in the proof of Theorem 4.2). This is one of the main reasons why we cannot prove more detailed results concerning the line index, such as for example its distribution. In particular, we believe that new tools are required to make progress on that front. This is also true for various related definitions to the line index as for instance the smallest number of vertices whose deletion results in a balanced graph and also the *feature line index* of $G_{n,m,p,q}$, namely the smallest number of feature choices whose inversion of signs results in a balanced graph. It is easy to verify that if the number of negative cycles is $\Omega(n^3)$, then the feature line index is $\Omega(n)$. However, it seems that it could be as large as $\Theta(nm)$. We leave the investigation of which bound is closer to the truth as an open problem.

Furthermore, to the best of our knowledge, research related to random signed graphs is still at an early stage. Taking into account the volume of results for random (unsigned) graphs and the special characteristics that signed graphs have, we believe that a new, quite interesting research area could be emerging. In particular, there are several important questions that remain open. For example, we mention the problem of determining the maximum size of a balanced clique in a signed graph, the existence of a balancing vertex (i.e. a vertex whose deletion along with all adjacent edges results in a balanced signed graph), the existence of negative cycles of size larger than 3, the existence of at most n negative faces provided that the underlying graph is planar, etc. Especially the last question is related to matroids since planar signed graphs with at most two vertex disjoint faces form the main building blocks for the class of quaternary and non-binary signed-graphic matroids, see e.g [13]. Moreover, it is known that a connected signed-graphic matroid is binary if and only if the associated signed graph has no two vertex disjoint negative cycles. Therefore, based on such statements, asymptotic results for signed graphs can be used in order to study the asymptotic behavior of the associated matroids. In view of this, we believe that research related to random signed graphs may also contribute to the recent development of a theory of matroid asymptotics (see e.g. [2, 9, 11]).

References

[1] N. Alon and J. H. Spencer, The Probabilistic Method. *John Wiley & Sons*, 2007

[2] N. Bansal, R. A. Pendavingh and J. G. van der Pol, On the number of matroids, *Combinatorica* 35 (2015), 253–277.

[3] M. Bloznelis, E. Godehardt, J. Jaworski, V. Kurauskas and K. Rybarczykl, Recent progress in complex network analysis — Models of random intersection graphs, In Data Science, Learning by Latent Structures, and Knowledge Discovery, *Springer*, 2015.

[4] D. Cartwright and F. Harary, Structural balance: A generalization of Heider's theory, *Psychological Review* 63 (1956), 277–293.

[5] F. Harary, On the notion of balance of a signed graph, *Michigan Math. J.* 2 (1954), 143–146.

[6] M. Karoński, E. R. Sheinerman and K. B. Singer-Cohen, On random intersection graphs: The subgraph problem, *Combin. Probab. Comput.* 8 (1999), 131–159.

[7] A. El Maftouhi, A. Harutyunyan and Y. Manoussakis, Weak balance in random signed graphs, *Internet Math.* 11 (2015), 143–154.

[8] A. El Maftouhi, Y. Manoussakis and O. Megalakaki, Balance in random signed graphs, *Internet Math.* 8 (2012), 364–380.

[9] D. Mayhew, M. Newman, D. Welsh and G. Whittle, On the asymptotic proportion of connected matroids, *European J. Combin.* 32 (2011), 882–890.

[10] S. E. Nikoletseas, C. L. Raptopoulos and P. G. Spirakis, Large independent sets in general random intersection graphs, *Theoret. Comput. Sci.* 406 (2008), 215–224.

[11] R. A. Pendavingh and J. G. van der Pol, On the number of matroids compared to the number of sparse paving matroids, *Electron. J. Combin.* 22(2) (2015), #P2.51.

[12] S. M. Ross, *A First Course in Probability*, Pearson, 2012.

[13] D. Slilaty and H. Qin, Decompositions of signed-graphic matroids, *Discrete Math.* 307 (2007), 2187–2199.

[14] T. Zaslavsky, A mathematical bibliography of signed and gain graphs and allied areas, *Electron. J. Combin.* #DS8 (2012).

(Received 9 Oct 2017; revised 21 Apr 2019)