CORRELATIONS FOR PATHS IN RANDOM ORIENTATIONS OF $G(n,p)$ AND $G(n,m)$

SVEN ERICK ALM, SVANTE JANSON, AND SVANTE LINUSSON

Abstract. We study random graphs, both $G(n,p)$ and $G(n,m)$, with random orientations on the edges. For three fixed distinct vertices s,a,b we study the correlation, in the combined probability space, of the events $\{a \to s\}$ and $\{s \to b\}$.

For $G(n,p)$, we prove that there is a $p_c = 1/2$ such that for a fixed $p < p_c$ the correlation is negative for large enough n and for $p > p_c$ the correlation is positive for large enough n. We conjecture that for a fixed $n \geq 27$ the correlation changes sign three times for three critical values of p.

For $G(n,m)$ it is similarly proved that, with $p = m/\binom{n}{2}$, there is a critical p_c that is the solution to a certain equation and approximately equal to 0.7993. A lemma, which computes the probability of non existence of any ℓ directed edges in $G(n,m)$, is thought to be of independent interest.

We present exact recursions to compute $\mathbb{P}(a \to s)$ and $\mathbb{P}(a \to s, s \to b)$. We also briefly discuss the corresponding question in the quenched version of the problem.

1. Introduction

For a graph $G = (V,E)$ we orient each edge with equal probability for the two possible directions and independent of all other edges. This model has been studied previously in for instance [1, 6, 12]. Let s,a,b be fixed distinct vertices in the graph. Let $\{a \to s\}$ denote the event that there exists a directed path from a to s. The object of this paper is to study the correlation of the two events $\{a \to s\}$ and $\{s \to b\}$ in a random graph. One might intuitively guess that they are negatively correlated in any graph, that is $\mathbb{P}(a \to s, s \to b) - \mathbb{P}(a \to s)\,\mathbb{P}(s \to b) < 0$. This is however not true for all graphs. In fact, the smallest counter example is the graph on four vertices with all edges except $\{a,b\}$. In [1] it was proved that for the complete graph K_n they are negatively correlated for $n = 3$, independent for $n = 4$ and positively correlated for $n \geq 5$. The results in [1] suggested that the correlation seemed more likely to be positive in dense graphs, which led us to further investigate the correlation in $G(n,p)$ and in $G(n,m)$ as presented in the present paper.

There are two different ways of combining the two random processes. In this paper we deal with the annealed case, which means that we consider the joined probability space of $G(n,p)$ (or $G(n,m)$) and that of edge orientations. We then consider correlations of events in this space. The other possibility is the quenched version, in which one for each graph in $G(n,p)$ (or $G(n,m)$) computes the correlation of the two events in the

Date: 8 June, 2010.

Svante Linusson is a Royal Swedish Academy of Sciences Research Fellow supported by a grant from the Knut and Alice Wallenberg Foundation.

This research was conducted when the authors visited Institut Mittag-Leffler (Djursholm, Sweden).

probability space of all edge orientations in that graph. Then one takes the expected value over all graphs. We discuss briefly the quenched version in Section 9.

In the random graph $G(n,p)$ there are n vertices and every edge exists with probability p, $0 \leq p \leq 1$, independently of other edges. Let $\vec{G}(n,p)$ be a directed graph obtained by giving each edge in $G(n,p)$ a random direction. Our main result in Section 2 implies that for a fixed p the events are positively correlated when $p > 1/2$ for large enough n, but negatively correlated when $p < 1/2$ for large enough n. To be more precise, in Theorem 2.1 we prove that

$$\frac{\mathbb{P}(a \not\to s, s \not\to b) - \mathbb{P}(a \not\to s)\,\mathbb{P}(s \not\to b)}{\mathbb{P}(a \not\to s, s \not\to b)} \to \frac{2p-1}{3} \text{ as } n \to \infty.$$

Note that the covariance is bilinear so that $\mathbb{P}(a \to s, s \to b) - \mathbb{P}(a \to s)\,\mathbb{P}(s \to b) = \mathbb{P}(a \not\to s, s \not\to b) - \mathbb{P}(a \not\to s)\,\mathbb{P}(s \not\to b)$. Thus we may instead study the complementary events $\{a \not\to s\}$, that there does not exist a directed path from a to s, and $\{s \not\to b\}$, which turns out to be more practical. The change of sign for the correlation at $p = 1/2$ was unexpected for us. An attempt to an heuristic explanation is given just after the statement of Theorem 2.1.

In Section 5 we give exact recursions for the probabilities $\mathbb{P}(a \not\to s), \mathbb{P}(s \not\to b)$ and $\mathbb{P}(a \not\to s, s \not\to b)$. This is done by studying the size of the set of the vertices that can be reached along a directed path from a given vertex s and the size of the set of vertices from which there is a directed path to s. We give recursions for the joint distribution of these sizes, which could be of independent interest.

The situation seems however to be much more complicated than what is suggested by Theorem 2.1. In Section 6 we report our results from exact computations using the recursions. It turns out that for $n \geq 27$ the difference $\mathbb{P}(a \to s, s \to b) - \mathbb{P}(a \to s) \cdot \mathbb{P}(s \to b)$ seems to change sign three times. First twice for small p (roughly *constant*$/n$) and then it goes from negative to positive again just before $p = 1/2$. We find this behavior mysterious but conjecture it to be true in general and plan to study this further in a coming paper.

For $0 \leq m \leq \binom{n}{2}$, $G(n,m)$ is the random graph with n vertices (which we take as $1, \dots, n$) and m edges, uniformly chosen among all such graphs. We let $\vec{G}(n,m)$ be the directed random graph obtained from $G(n,m)$ by giving each edge a random direction. It is less straightforward to study the probabilities in $\vec{G}(n,m)$, but thanks to Lemma 3.2, we can prove the corresponding statement for $\vec{G}(n,m)$ in parallel with the proof for $\vec{G}(n,p)$. Here the critical probability is ≈ 0.799288221, see Theorem 2.2. The exact computations and the simulations for $\vec{G}(n,m)$ do not give reason to believe that there is more than one critical probability for a fixed n. It might be consternating that $\vec{G}(n,m)$ and $\vec{G}(n,p)$ behave so differently. In Section 8 we explain this and prove that the correlation in $\vec{G}(n,p)$ is always larger, in a certain sense, than in $\vec{G}(n,m)$. The difference is expressed as a variance for which we give an exact asymptotic expression, see Theorem 8.2.

The background of the questions studied in this paper is as follows. Let G be any graph and a, b, s, t distinct vertices of G. Further, assign a random orientation to the graph as described above, that is each edge is given its direction independent of the other edges. In [12] it was proved that then the events $\{s \to a\}$ and $\{s \to b\}$ are positively correlated. This was shown to be true also if we first conditioned on $\{s \not\to t\}$, i.e. $\mathbb{P}(s \to a, s \to b \mid s \not\to t) \geq \mathbb{P}(s \to a \mid s \not\to t) \cdot \mathbb{P}(s \to b \mid s \not\to t)$. Note that it is not intuitively clear why this is so. It is for instance no longer true if we instead condition on $\{s \to t\}$. In another direction, it was proved that $\mathbb{P}(s \to b, a \to t \mid s \not\to t) \leq \mathbb{P}(s \to b \mid s \not\to t) \cdot \mathbb{P}(a \to t \mid s \not\to t)$. The proofs in [12] relied heavily on the results in [5] and [4] where similar statements were proved for edge percolation on a given graph. These questions have to a large extent been inspired by an interesting conjecture due to Kasteleyn named the Bunkbed conjecture by Häggström [8], see also [11] and Remark 5 in [5].

Acknowledgment: We thank the anonymous referees for very helpful comments.

2. Main Theorems

Let $G(n,p)$ be the random graph where each edge has probability p of being present independent of other edges. We further orient each present edge either way independently with probability $\frac{1}{2}$ and call the resulting random directed graph $\vec{G}(n,p)$. It turns out to be convenient to study the negated events $\{a \not\to s\}$ and $\{s \not\to b\}$. We say that two events A and B are negatively correlated if

$$\mathrm{Cov}(A,B) := \mathbb{P}(A \cap B) - \mathbb{P}(A)\,\mathbb{P}(B) < 0,$$

and positively correlated if the inequality goes the other way. If A and B are negatively correlated, then A and $\neg B$ are positively correlated. Thus $a \not\to s$ and $s \not\to b$ are negatively correlated if and only if $\{a \to s\}$ and $\{s \to b\}$ are negatively correlated. Trivially, $\mathbb{P}(a \not\to s) = \mathbb{P}(s \not\to b)$.

All unspecified limits are as $n \to \infty$.

The case $p = 1$ corresponds to random orientations of the complete graph K_n. In [1] it was proved that $a \not\to s$ and $s \not\to b$ are negatively correlated in K_3, independent in K_4 and positively correlated in K_n for $n \geq 5$. It was in fact proved that the relative covariance $(\mathbb{P}(a \not\to s, s \not\to b) - \mathbb{P}(a \not\to s)\,\mathbb{P}(s \not\to b))/\,\mathbb{P}(a \not\to s, s \not\to b)$ converged to $\frac{1}{3}$ as $n \to \infty$.

For $\vec{G}(n,p)$ our main theorem is as follows.

Theorem 2.1. *For a fixed $p \in (0,1]$ and fixed distinct vertices $a, b, s \in \vec{G}(n,p)$ we have the following limit of the relative covariance*

$$\lim_{n\to\infty} \frac{\mathbb{P}(a \not\to s, s \not\to b) - \mathbb{P}(a \not\to s)\,\mathbb{P}(s \not\to b)}{\mathbb{P}(a \not\to s, s \not\to b)} = \frac{2p-1}{3}.$$

In particular, for large enough n:

(i) *If $0 < p < 1/2$, then* $\mathrm{Cov}(s \not\to a,\, b \not\to s) < 0$.
(ii) *If $1/2 < p \leq 1$, then* $\mathrm{Cov}(s \not\to a,\, b \not\to s) > 0$.

Furthermore,

$$\mathrm{Cov}(s \not\to a,\, b \not\to s) = (2p - 1 + o(1))(1 - p/2)^{2n-3}. \tag{1}$$

One way to heuristically understand the critical probability 1/2 is as follows. For the event $\{s \not\to a\}$ the two cases that dominate for large n are $i)$ no edge points out from s and $ii)$ no edge points in to a, each with probability $(1-p/2)^{n-1}$. Similarly for $\{b \not\to s\}$. Thus $\mathbb{P}(s \not\to a)\,\mathbb{P}(b \not\to s)$ is roughly $4 \cdot (1-p/2)^{2n-2}$. For the joint probability $\mathbb{P}(s \not\to a, b \not\to s)$ on the other hand, we only need to consider three of these cases since no edges at all around s has a much smaller probability. This gives a negative contribution to the covariance and is perhaps the most striking contribution. But, the other three cases have in the joint probability a factor $1-p/2$ less since one edge is common, which gives a positive contribution. Thus the quotient $\frac{\mathbb{P}(s\not\to a, b\not\to s)}{\mathbb{P}(s\not\to a)\,\mathbb{P}(b\not\to s)} \sim \frac{3/4}{1-p/2}$, which is greater than 1 if and only if $p > 1/2$.

Theorem 2.1 was proved in an earlier version of this manuscript, see [2], with a rather lengthy calculation to estimate the involved probabilities. In the present version we will present a shorter proof which also has the advantage that it allows us to prove the covariance also in $\vec{G}(n,m)$. However, we want to point out that the lengthy calculations in [2] could be shortened with a clever lemma by Backelin, see [3].

For $\vec{G}(n,m)$, we obtain the corresponding result.

Theorem 2.2. *Suppose that $m,n \to \infty$, and that $m/\binom{n}{2} \to p \in (0,1]$. Let $p_{\mathsf{c}} \approx 0.799$ be the unique root in $[0,1]$ of $3e^{-\frac{2p(1-p)}{(2-p)^2}} = 4-2p$. If n is large enough, then for any distinct vertices $s,a,b \in \vec{G}(n,m)$:*

(i) *If $0 < p < p_{\mathsf{c}}$, then $\operatorname{Cov}(s \not\to a,\ b \not\to s) < 0$.*
(ii) *If $p_{\mathsf{c}} < p \le 1$, then $\operatorname{Cov}(s \not\to a,\ b \not\to s) > 0$.*

In fact we have the following asymptotic formula for the covariance.

Theorem 2.3. *Suppose that $m,n \to \infty$, let $p = p(n) := m/\binom{n}{2}$, and assume $\liminf p(n) > 0$. Then, for any distinct vertices $s,a,b \in \vec{G}(n,m)$:*

$$\begin{aligned}\operatorname{Cov}(s \not\to a,\ b \not\to s) &= \left(\frac{3}{4-2p}e^{-\frac{2p(1-p)}{(2-p)^2}} - 1 + o(1)\right)\mathbb{P}(s \not\to a)^2 \\ &= \left(3 \cdot e^{-\frac{2p(1-p)}{(2-p)^2}} - 4 + 2p + o(1)\right) \cdot (1-p/2)^{2n-3} \cdot e^{-\frac{2p(1-p)}{(2-p)^2}}.\end{aligned}$$

3. Two lemmas

In order to prove the theorems for $\vec{G}(n,p)$ and $\vec{G}(n,m)$ in parallel we need two lemmas.

Lemma 3.1. *Fix $\ell \ge 0$ edges in K_n. The probability that none of these ℓ edges appears in $G(n,m)$ is at most $\left(1 - m/\binom{n}{2}\right)^{\ell}$.*

Proof. Let $N := \binom{n}{2}$. We may assume that $m + \ell \le N$. Then the probability is

$$\frac{\binom{N-\ell}{m}}{\binom{N}{m}} = \frac{(N-\ell)!\,(N-m)!}{(N-\ell-m)!\,N!} = \prod_{i=0}^{\ell-1}\frac{N-m-i}{N-i} \le \prod_{i=0}^{\ell-1}\Bigl(1 - \frac{m}{N}\Bigr). \qquad \square$$

Note that in $G(n,p)$ the corresponding probability is exactly $(1-p)^{\ell}$ by independence of the edges. We next show a more precise version of Lemma 3.1 for the directed graph $\vec{G}(n,m)$.

Let $q(\ell) = q(\ell; n, m)$ be the probability that if we fix ℓ edges in K_n and give each an orientation, then none of these directed edges appears in $\vec{G}(n, m)$. (Here $0 \le \ell \le \binom{n}{2}$. By symmetry, the probability does not depend on the choice of the ℓ edges and their orientations.) Let also $q'(\ell; n, p)$ be the corresponding probability in $\vec{G}(n, p)$. It is clear that $q'(\ell; n, p) = (1 - p/2)^\ell$. For $\vec{G}(n, m)$ we have the following result, which we believe is of independent interest.

Lemma 3.2. *Suppose that $\ell = \ell(n) = O(n)$ and that $0 \le m = m(n) \le \binom{n}{2}$. Then, with $p = p(n) = m(n)/\binom{n}{2}$, as $n \to \infty$,*

$$q(\ell; n, m) \sim (1 - p/2)^\ell \exp\left(-\left(\frac{\ell}{n}\right)^2 \frac{p(1-p)}{(2-p)^2}\right). \tag{2}$$

Furthermore, for any ℓ, n, m we have $q(\ell; n, m) \le q'(\ell; n, p) = (1 - p/2)^\ell$.

Proof. Let $N := \binom{n}{2}$. The cases $m = 0$ and $m = N$ are trivial, so we may assume that $0 < m < N$ and thus $0 < p < 1$.

Let X be the number of the ℓ chosen edges that appear in $G(n, m)$, ignoring orientations. Then X has the hypergeometric distribution

$$\mathbb{P}(X = k) = \frac{\binom{\ell}{k}\binom{N-\ell}{m-k}}{\binom{N}{m}} = \binom{\ell}{k} \frac{(N-\ell)!\,(N-m)!\,m!}{N!\,(N-m-\ell+k)!\,(m-k)!}. \tag{3}$$

Define, for $J, j \ge 0$, $a(J, j) := J^{-j} J!/(J-j)!$ (interpreted as 0 when $j > J$). Then (3) can be written as

$$\begin{aligned}
\mathbb{P}(X = k) &= \binom{\ell}{k} a(N, \ell)^{-1} N^{-\ell} a(N-m, \ell-k)(N-m)^{\ell-k} a(m, k) m^k \\
&= \binom{\ell}{k} p^k (1-p)^{\ell-k} a(N, \ell)^{-1} a(N-m, \ell-k) a(m, k).
\end{aligned}$$

Given $X = k$, the probability that none of the selected directed edges appears in $\vec{G}(n, m)$ equals the probability that none of the k selected edges that appear in $G(n, m)$ gets the forbidden orientation, which equals $(1/2)^k$. Hence,

$$\begin{aligned}
q(\ell) &= \sum_{k=0}^{\ell} 2^{-k}\, \mathbb{P}(X = k) \\
&= \sum_{k=0}^{\ell} \binom{\ell}{k} \left(\frac{p}{2}\right)^k (1-p)^{\ell-k} a(N, \ell)^{-1} a(N-m, \ell-k) a(m, k) \\
&= a(N, \ell)^{-1} (1 - p/2)^\ell \sum_{k=0}^{\ell} \binom{\ell}{k} \left(\frac{p/2}{1-p/2}\right)^k \left(1 - \frac{p/2}{1-p/2}\right)^{\ell-k} a(N-m, \ell-k) a(m, k) \\
&= a(N, \ell)^{-1} (1 - p/2)^\ell\, \mathbb{E}\big(a(m, Y) a(N-m, \ell-Y)\big),
\end{aligned}$$

where Y has the binomial distribution $\mathrm{Bin}(\ell, p/(2-p))$. Consequently, (2) is equivalent to

$$a(N, \ell)^{-1}\, \mathbb{E}\big(a(m, Y) a(N-m, \ell-Y)\big) \sim \exp\left(-\left(\frac{\ell}{n}\right)^2 \frac{p(1-p)}{(2-p)^2}\right). \tag{4}$$

In order to show (4), it suffices to show that every subsequence has a subsequence where (4) holds. (See [10, p. 12].) Consequently, by choosing suitable subsequences, we may assume that

$$\ell/n \to \lambda \qquad \text{and} \qquad p \to \pi, \tag{5}$$

for some $\lambda \in [0, \infty)$ and $\pi \in [0, 1]$; furthermore, we may assume that either

$$\ell \to \infty \qquad \text{or} \qquad l = O(1). \tag{6}$$

We next observe that for all $J \geq 1$ and $j \leq J$,

$$a(J, j) = J^{-j} \prod_{i=0}^{j-1} (J - i) = \prod_{i=0}^{j-1} \Bigl(1 - \frac{i}{J}\Bigr) = \exp\left(\sum_{i=1}^{j-1} \log\Bigl(1 - \frac{i}{J}\Bigr)\right).$$

It follows that $0 \leq a(j, J) \leq 1$ for all j and J. Moreover, if $j = O(J^{1/2})$, then, if also $j \leq J$,

$$\begin{aligned} \sum_{i=1}^{j-1} \log\Bigl(1 - \frac{i}{J}\Bigr) &= \sum_{i=1}^{j-1} \Bigl(-\frac{i}{J} + O\Bigl(\frac{i}{J}\Bigr)^2\Bigr) = -\frac{j^2 - j}{2J} + O\Bigl(\frac{j^3}{J^2}\Bigr) \\ &= -\frac{j^2}{2J} + O\bigl(J^{-1/2}\bigr) \end{aligned}$$

and thus (also in the trivial case $j > J$)

$$a(J, j) = \exp\Bigl(-\frac{j^2}{2J}\Bigr) + O\bigl(J^{-1/2}\bigr). \tag{7}$$

In particular, (7) applies to $a(N, \ell)$, because $\ell = O(n) = O(N^{1/2})$, and thus

$$a(N, \ell) = \exp\bigl(-\ell^2/2N\bigr) + o(1) \to \exp\bigl(-\lambda^2\bigr), \tag{8}$$

assuming (5).

Next, consider $a(m, Y)$, assuming (5)–(6). Assume first $\pi > 0$. If also $\ell \to \infty$, then by the law of large numbers, $Y/\ell \xrightarrow{\mathrm{p}} \pi/(2 - \pi)$, and thus

$$\frac{Y^2}{2m} = \Bigl(\frac{Y}{\ell}\Bigr)^2 \Bigl(\frac{\ell}{n}\Bigr)^2 \cdot \frac{n^2/2}{m} \xrightarrow{\mathrm{p}} \Bigl(\frac{\pi}{2 - \pi}\Bigr)^2 \lambda^2 \frac{1}{\pi} = \frac{\pi}{(2 - \pi)^2} \lambda^2. \tag{9}$$

On the other hand, if $\ell = O(1)$, then $\lambda = 0$; furthermore, $Y \leq \ell = O(1)$ and $m \to \infty$ (because $\pi > 0$), so $Y^2/2m \xrightarrow{\mathrm{p}} 0$; hence (9) holds in this case too.

In the case $\pi = 0$, so $p \to 0$, we instead use

$$\mathbb{E}\, Y = \ell p/(2 - p) \leq p\ell = o\bigl(p^{1/2}\ell\bigr),$$

and thus $Y/(p^{1/2}\ell) \xrightarrow{\mathrm{p}} 0$. Hence,

$$\frac{Y^2}{2m} = \left(\frac{Y}{p^{1/2}\ell}\right)^2 \cdot \frac{p\ell^2}{2m} = \left(\frac{Y}{p^{1/2}\ell}\right)^2 \cdot \frac{\ell^2}{n(n-1)} \xrightarrow{\mathrm{p}} 0 = \frac{\pi}{(2 - \pi)^2} \lambda^2,$$

so (9) holds in this case too.

Consequently, (9) holds in every case. In particular, $Y = O_{\mathrm{p}}(m^{1/2})$, and (7) implies that, provided $m \to \infty$,

$$a(m, Y) = \exp\left(-\frac{Y^2}{2m}\right) + o_{\mathrm{p}}(1) \xrightarrow{\mathrm{p}} \exp\left(-\frac{\pi}{(2 - \pi)^2} \lambda^2\right); \tag{10}$$

on the other hand, if $m = O(1)$, then $\mathbb{P}(Y = 0) \to 1$, e.g. by (9), so (10) holds in this case too. Consequently, (10) holds in every case.

Since $\ell - Y \sim \mathrm{Bin}\big(\ell, 1 - p/(2-p)\big) = \mathrm{Bin}(\ell, 2(1-p)/(2-p))$, it follows by similar arguments, now considering the case $\pi = 1$ separately, that

$$\frac{(\ell - Y)^2}{2(N-m)} \xrightarrow{\text{p}} \frac{4(1-\pi)}{(2-\pi)^2}\lambda^2$$

and

$$a(N-m, \ell - Y) \xrightarrow{\text{p}} \exp\left(-\frac{4(1-\pi)}{(2-\pi)^2}\lambda^2\right). \tag{11}$$

Combining (10) and (11) we find

$$a(m,Y)a(N-m,\ell-Y) \xrightarrow{\text{p}} \exp\left(-\frac{\lambda^2}{(2-\pi)^2}\big(\pi + 4(1-\pi)\big)\right).$$

By dominated convergence, using $0 \le a(m,Y)a(N-m,\ell-Y) \le 1$, this, together with (8), yields

$$\begin{aligned} a(N,\ell)^{-1}\,\mathbb{E}\big(a(m,Y)a(N-m,\ell-Y)\big) &\to \exp\left(\lambda^2 - \frac{\lambda^2}{(2-\pi)^2}(4-3\pi)\right) \\ &= \exp\left(\frac{\lambda^2}{(2-\pi)^2}(\pi^2 - \pi)\right), \end{aligned}$$

which is equivalent to (4) by (5). Hence (4) holds assuming (5)–(6), and thus in general, which shows (2).

Finally, let $X' \sim \mathrm{Bin}(\ell, p)$ be the number of the ℓ chosen edges that appear in $G(n,p)$, ignoring orientations. Then, $q(\ell) = \mathbb{E}\,(\frac{1}{2})^X$ as shown above, and similarly $q'(\ell) = \mathbb{E}\,(\frac{1}{2})^{X'}$. In fact, $\mathbb{E}\,a^X \le \mathbb{E}\,a^{X'}$ for any $a > 0$, see Hoeffding [9, Theorem 4], which, taking $a = 1/2$, yields $q(\ell) \le q'(\ell)$. □

4. Proofs

Say that a set $S \subseteq V(G)$ of vertices in a directed graph G is an *outset* if there is no directed edge vw with $v \notin S$, $w \in S$ (i.e., all edges between S and its complement are directed out from S), and an *inset* if there is no directed edge vw with $v \in S$, $w \notin S$. Hence, S is an outset if and only if its complement is an inset.

Lemma 4.1. *If G is a directed graph and $a, b \in V(G)$, then the following are equivalent:*
- (i) $a \not\to b$.
- (ii) *There exists an inset S with $a \in S$, $b \notin S$.*
- (iii) *There exists an outset T with $a \notin T$, $b \in T$.*

Proof. (i) $\iff$ (ii): No directed path can leave an inset. Hence, if an inset S as in (ii) exists, then $a \not\to b$. Conversely, if $a \not\to b$, then $S := \{v \in V(G) : a \to v\}$ is an inset with $a \in S$, $b \notin S$.

(ii) $\iff$ (iii): Take $T := V(G) \setminus S$, and conversely. □

For $S \subseteq [n]$, let $\mathcal{O}_S$ and $\mathcal{I}_S$ denote the events $\{S \text{ is an outset}\}$ and $\{S \text{ is an inset}\}$, respectively. The events will be in $\vec{G}(n,m)$ or $\vec{G}(n,p)$ depending on context. We also write, for typographical convenience, $\mathcal{O}_a := \mathcal{O}_{\{a\}}$, etc.

Lemma 4.2. *Assume that $p \in (0,1]$ is fixed. Then, for any distinct $s, a \in \vec{G}(n,p)$,*

$$\begin{aligned}\mathbb{P}(s \not\to a) &= \mathbb{P}(\mathcal{I}_s) + \mathbb{P}(\mathcal{O}_a) + O\big(n(1-p/2)^{2n}\big)\\ &= 2q'(n-1) + O\big(n(1-p/2)^{2n}\big)\\ &\sim 2(1-p/2)^{n-1}.\end{aligned}$$

The proof of Lemma 4.2 is very similar to the proof of Lemma 4.3 below, but with $q(\ell; n, m)$ replaced by $q'(\ell; n, p)$.

Lemma 4.3. *Suppose that $m, n \to \infty$, let $p = p(n) := m/\binom{n}{2}$, and assume $\liminf p(n) > 0$. Then, for any distinct $s, a \in \vec{G}(n,m)$,*

$$\begin{aligned}\mathbb{P}(s \not\to a) &= \mathbb{P}(\mathcal{I}_s) + \mathbb{P}(\mathcal{O}_a) + O\big(n(1-p/2)^{2n}\big)\\ &= 2q(n-1) + O\big(n(1-p/2)^{2n}\big)\\ &\sim 2(1-p/2)^{n-1}e^{-\frac{p(1-p)}{(2-p)^2}}.\end{aligned}$$

Proof. Let $\mathcal{E}_k$ be the event that there exists an inset S with $s \in S$, $a \notin S$ and $|S| = k$. Then, by Lemma 4.1,

$$\{s \not\to a\} = \bigcup_{k=1}^{n-1} \mathcal{E}_k = \mathcal{E}_1 \cup \mathcal{E}_{n-1} \cup \bigcup_{k=2}^{n-2} \mathcal{E}_k. \tag{12}$$

Here $\mathcal{E}_1 = \mathcal{I}_s$ and $\mathcal{E}_{n-1} = \mathcal{O}_a$.

If $|S| = k$, then $\mathbb{P}(\mathcal{I}_S) = \mathbb{P}(\mathcal{O}_S) = q(k(n-k))$. In particular, by Lemma 3.2,

$$\mathbb{P}(\mathcal{I}_s) = \mathbb{P}(\mathcal{O}_a) = q(n-1) \sim (1-p/2)^{n-1}e^{-\frac{p(1-p)}{(2-p)^2}}. \tag{13}$$

Further, for any k, there are $\binom{n-2}{k-1}$ sets S with $s \in S$, $a \notin S$ and $|S| = k$, and thus

$$\mathbb{P}(\mathcal{E}_k) \le \binom{n-2}{k-1} q(k(n-k)) \le n^{k-1} q(k(n-k)). \tag{14}$$

For fixed k we may apply Lemma 3.2 and obtain

$$\mathbb{P}(\mathcal{E}_k) = \mathbb{P}(\mathcal{E}_{n-k}) = O\big(n^{k-1}(1-p/2)^{k(n-k)}\big) = O\big(n^{k-1}(1-p/2)^{kn}\big). \tag{15}$$

For a fixed K, we use this estimate for $k < K$. For $K \le k \le n-K$ we have $k(n-k) \ge K(n-K)$ and thus $q(k(n-k)) \le q(K(n-K))$, while the total number of subsets S of $[n]$ is 2^n, and thus Lemma 3.2 again yields

$$\mathbb{P}\Big(\bigcup_{k=K}^{n-K} \mathcal{E}_k\Big) \le 2^n q(K(n-K)) = O\big(2^n(1-p/2)^{Kn}\big). \tag{16}$$

By the assumption $\liminf p(n) > 0$, we may assume that $p = p(n) \ge p_0$ for some $p_0 > 0$. Choosing K such that $(1-p_0/2)^{K-2} \le 1/2$, we have $2(1-p/2)^K \le 2(1-p_0/2)^{K-2}(1-p/2)^2 \le (1-p/2)^2$, and thus (15) and (16) imply

$$\mathbb{P}\Big(\bigcup_{k=2}^{n-2} \mathcal{E}_k\Big) = \sum_{k=2}^{K-1} O\big(n^{k-1}(1-p/2)^{kn}\big) + O\big((1-p/2)^{2n}\big) = O\big(n(1-p/2)^{2n}\big).$$

Hence, (12) yields

$$\mathbb{P}\{s \not\to a\} = \mathbb{P}(\mathcal{E}_1 \cup \mathcal{E}_{n-1}) + O\big(n(1-p/2)^{2n}\big) = \mathbb{P}(\mathcal{I}_s \cup \mathcal{O}_a) + O\big(n(1-p/2)^{2n}\big). \quad (17)$$

Finally, $\mathbb{P}(\mathcal{I}_s \cap \mathcal{O}_a) = q(2n-3) = O\big((1-p/2)^{2n}\big)$ by Lemma 3.2 again, and thus $\mathbb{P}(\mathcal{I}_s \cup \mathcal{O}_a) = \mathbb{P}(\mathcal{I}_s) + \mathbb{P}(\mathcal{O}_a) + O\big((1-p/2)^{2n}\big)$. The result now follows by (17) and (13). ☐

Lemma 4.4. *Suppose that $p \in (0,1]$ is fixed, then for any distinct $s, a, b \in \vec{G}(n,p)$,*

$$\begin{aligned}\mathbb{P}(s \not\to a \text{ and } b \not\to s) & \\ &= \mathbb{P}(\mathcal{I}_{s,b} \cap \mathcal{I}_b) + \mathbb{P}(\mathcal{O}_{s,a} \cap \mathcal{O}_a) + \mathbb{P}(\mathcal{O}_a \cap \mathcal{I}_b) + o\big((1-p/2)^{2n}\big) \\ &= 3(1-p/2)^{2n-3} + o\big((1-p/2)^{2n}\big)\end{aligned}$$

The proof of Lemma 4.4 is almost identical to that of Lemma 4.5 below.

Lemma 4.5. *Suppose that $m, n \to \infty$, let $p = p(n) := m/\binom{n}{2}$, and assume* $\liminf p(n) >$ 0*. Then, for any distinct $s, a, b \in [n]$,*

$$\begin{aligned}\mathbb{P}(s \not\to a \text{ and } b \not\to s) & \\ &= \mathbb{P}(\mathcal{I}_{s,b} \cap \mathcal{I}_b) + \mathbb{P}(\mathcal{O}_{s,a} \cap \mathcal{O}_a) + \mathbb{P}(\mathcal{O}_a \cap \mathcal{I}_b) + o\big((1-p/2)^{2n}\big) \\ &= 3q(2n-3) + o\big((1-p/2)^{2n}\big) \\ &\sim 3(1-p/2)^{2n-3} e^{-4\frac{p(1-p)}{(2-p)^2}}.\end{aligned}$$

Proof. As in (12), we have

$$\{s \not\to a\} = \bigcup_{k=1}^{n-1} \mathcal{E}_k \quad \text{with} \quad \mathcal{E}_k := \bigcup_{S: s\in S,\, a\notin S,\, |S|=k} \mathcal{I}_S.$$

and similarly

$$\{b \not\to s\} = \bigcup_{k=1}^{n-1} \mathcal{E}'_k \quad \text{with} \quad \mathcal{E}'_k := \bigcup_{S: s\in S,\, b\notin S,\, |S|=k} \mathcal{O}_S.$$

First, by the argument in the proof of Lemma 4.3 (choosing K large enough),

$$\mathbb{P}\Big(\bigcup_{k=3}^{n-3} \mathcal{E}_k\Big) = \mathbb{P}\Big(\bigcup_{k=3}^{n-3} \mathcal{E}'_k\Big) = O\big(n^2(1-p/2)^{3n}\big) = o\big((1-p/2)^{2n}\big).$$

We may thus approximate the event $\{s \not\to a \text{ and } b \not\to s\}$ by

$$\begin{aligned}(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_{n-2} \cup \mathcal{E}_{n-1}) \cap (\mathcal{E}'_1 \cup \mathcal{E}'_2 \cup \mathcal{E}'_{n-2} \cup \mathcal{E}'_{n-1}) & \\ &= \Bigg(\bigcup_{x\neq a} \mathcal{I}_{s,x} \cup \bigcup_{y\neq s} \mathcal{O}_{a,y}\Bigg) \cap \Bigg(\bigcup_{z\neq b} \mathcal{O}_{s,z} \cup \bigcup_{w\neq s} \mathcal{I}_{b,w}\Bigg)\end{aligned}$$

(taking notational advantage of $I_s = I_{s,s}$ and so on). This can be expanded as the union of $O(n^2)$ events of the type $\mathcal{X}_{S_1} \cap \mathcal{Y}_{S_2}$, where $\mathcal{X}$ and $\mathcal{Y}$ are $\mathcal{I}$ or $\mathcal{O}$, and $1 \le |S_1|, |S_2| \le 2$.

If $|S_1 \cup S_2| \geq 3$, then the event $\mathcal{X}_{S_1} \cap \mathcal{Y}_{S_2}$ forbids at least $|S_1 \cup S_2|(n - |S_1 \cup S_2|) \geq 3(n-4)$ directed edges, and Lemma 3.2 shows that

$$\mathbb{P}(\mathcal{X}_{S_1} \cap \mathcal{Y}_{S_2}) = O\big((1-p/2)^{3n}\big); \tag{18}$$

thus all such combinations have together a probability $O\big(n^2(1-p/2)^{3n}\big) = o\big((1-p/2)^{2n}\big)$ and may be ignored.

Furthermore, any event $\mathcal{I}_{s,x} \cap \mathcal{O}_{s,z}$ forbids at least $n-3$ edges (regardless of orientation) in $G(n,m)$, and thus, by Lemma 3.1,

$$\mathbb{P}(\mathcal{I}_{s,x} \cap \mathcal{O}_{s,z}) \leq (1-p)^{n-3}.$$

Since $p \geq p_0 > 0$, we have

$$\frac{(1-p/2)^2}{1-p} = \frac{1-p+p^2/4}{1-p} \geq 1 + p_0^2/4,$$

and thus $n^2(1-p)^{n-3} = o\big((1-p/2)^{2(n-3)}\big) = o\big((1-p/2)^{2n}\big)$, so may also ignore all $\mathcal{I}_{s,x} \cap \mathcal{O}_{s,z}$. We may similarly ignore $\mathcal{O}_a \cap \mathcal{I}_{b,a}$, $\mathcal{O}_{a,b} \cap \mathcal{I}_b$ and $\mathcal{O}_{a,b} \cap \mathcal{I}_{b,a}$.

This leaves only the events $\mathcal{I}_{s,b} \cap \mathcal{I}_b$, $\mathcal{I}_s \cap \mathcal{I}_b$, $\mathcal{O}_a \cap \mathcal{I}_b$, $\mathcal{O}_a \cap \mathcal{O}_s$, $\mathcal{O}_a \cap \mathcal{O}_{s,a}$. Of these, $\mathcal{I}_s \cap \mathcal{I}_b \subseteq \mathcal{I}_{s,b} \cap \mathcal{I}_b$ and $\mathcal{O}_a \cap \mathcal{O}_s \subseteq \mathcal{O}_a \cap \mathcal{O}_{s,a}$. Summarizing, we have found

$$\mathbb{P}(s \not\to a \text{ and } b \not\to s) = \mathbb{P}\big((\mathcal{I}_{s,b} \cap \mathcal{I}_b) \cup (\mathcal{O}_{s,a} \cap \mathcal{O}_a) \cup (\mathcal{O}_a \cap \mathcal{I}_b)\big) + o\big((1-p/2)^{2n}\big).$$

The intersection of any two of the events $\mathcal{I}_{s,b} \cap \mathcal{I}_b$, $\mathcal{O}_{s,a} \cap \mathcal{O}_a$ and $\mathcal{O}_a \cap \mathcal{I}_b$ is contained in an event $\mathcal{X}_{S_1} \cap \mathcal{Y}_{S_2}$ with $|S_1 \cup S_2| = 3$, so by (18), its probability is $o\big((1-p/2)^{2n}\big)$. Hence the first equality in the statement follows.

Moreover, each of these three events forbids exactly $2n-3$ directed edges, and thus each has the probability $q(2n-3)$. The result follows by Lemma 3.2. ☐

Note the obvious identity $\mathbb{P}(s \not\to a) = \mathbb{P}(b \not\to s)$.

Proof of Theorem 2.1. Follows easily from Lemmas 4.2 and 4.4. ☐

Proof of Theorem 2.3. An immediate consequence of Lemmas 4.3 and 4.5. ☐

Proof of Theorem 2.2. Let $f(p) := 3e^{-\frac{2p(1-p)}{(2-p)^2}} - 4 + 2p$. Note that $f(0) = -1$ and $f(1) = 1$. We will show that $f'(p) > 0$ for $p \in [0,1]$; this implies the existence of a unique zero $p_{\mathsf{c}} \in [0,1]$ of f, with $f(p) < 0$ for $0 \leq p < p_{\mathsf{c}}$ and $f(p) > 0$ for $p_{\mathsf{c}} < p \leq 1$. The result now follows from Theorem 2.3. The numerical value $p_{\mathsf{c}} \approx 0.799288221$ is found by `Maple`.

We use the partial fraction expansion $-p(1-p)/(2-p)^2 = 1 - 3(2-p)^{-1} + 2(2-p)^{-2}$ to find

$$f'(p) = 6e^{-\frac{2p(1-p)}{(2-p)^2}}\Big(-\frac{3}{(2-p)^2} + \frac{4}{(2-p)^3}\Big) + 2.$$

The factor $-3/(2-p)^2 + 4/(2-p)^3$ is increasing on $[0,1]$, as shown by another differentiation, and is thus at least $-3/4 + 4/8 = -1/4$. Since $0 < e^{-\frac{2p(1-p)}{(2-p)^2}} \leq 1$, it follows that $f'(p) \geq -6/4 + 2 = 1/2 > 0$ for $p \in [0,1]$. ☐

Remark 4.6. *In the proofs of Lemmas 4.2 and 4.4, all error terms are exponentially small compared to the main terms. Consequently,* $\operatorname{Cov}(s \not\to a, b \not\to s)/(1-p/2)^{2n-3}$ *approaches its limit* $2p-1$ *exponentially fast for any fixed* $p > 0$ *(this further holds* *uniformly for $p \geq p_0$ for any $p_0 > 0$), and similarly, the relative covariance approaches its limit in Theorem 2.1 exponentially fast for any fixed $p > 0$.*

Note, however, that (1) *is false in the trivial case $p = 0$, which shows that the convergence cannot be uniform for all $p > 0$.*

Problem 4.7. *It would be interesting to know the rate of convergence in $G(n, m)$.*

5. Exact recursions in $\vec{G}(n,p)$

In this section we will give exact recursions to compute

$$f_n(p) := \mathbb{P}_{\vec{G}(n,p)}(s \not\to b) \text{ and } g_n(p) := \mathbb{P}_{\vec{G}(n,p)}(a \not\to s, s \not\to b).$$

For a vertex $v \in V(G)$, let $\overrightarrow{C}_v \subseteq V(G)$ be the (random) set of all vertices u for which there is a directed path from v to u. We will call this the **out-cluster** from v. Let also the **in-cluster**, $\overleftarrow{C}_v \subseteq V(G)$ be the (random) set of all vertices u for which there is a directed path from u to v. Note that we will use the convention that $v \in \overleftarrow{C}_v \cap \overrightarrow{C}_v$. Let $y := 1 - p/2$ be the probability that an edge does not exist with a certain direction, and let $q := 1 - p$ be the probability that there is no edge at all.

For $n \geq 1$, $s \in X \subseteq [n]$ and $|X| = k$ define:

$$d_p(n,k) := \mathbb{P}_{\vec{G}(n,p)}(\overrightarrow{C}_s = X),$$

where in particular $d_p(1,1) = 1$.

Lemma 5.1. *We have the following recursions*

(i) $d_p(n,k) = d_p(k,k)y^{k(n-k)}$, *for $n > k \geq 1$,*

(ii) $d_p(k,k) = 1 - \sum_{i=1}^{k-1} \binom{k-1}{i-1} d_p(i,i) y^{i(k-i)}$, *and*

(iii) $f_n(p) = \sum_{k=1}^{n-1} \binom{n-2}{k-1} d_p(k,k) y^{k(n-k)}$.

Proof. If $n > k$ there is a vertex $w \notin X$. The only restriction on w is that it must not have any edge directed from X so $\mathbb{P}_{\vec{G}(n,p)}(\overrightarrow{C}_s = X) = \mathbb{P}_{G(n-1,p)}(\overrightarrow{C}_s = X) \cdot y^k$ and (i) follows by induction. Clearly $\sum_{X: s \in X \subseteq [n]} \mathbb{P}_{\vec{G}(n,p)}(\overrightarrow{C}_s = X) = 1$, which gives formula (ii) after using equation (i). To get the third equation we sum over all sets X that contain s but not b, i.e. $f_n(p) = \sum_X \mathbb{P}_{\vec{G}(n,p)}(\overrightarrow{C}_s = X) = \sum_{k=1}^{n-1} \binom{n-2}{k-1} d_p(n,k)$, which using (i) gives (iii). □

Note that, by symmetry, also $\mathbb{P}_{\vec{G}(n,p)}(\overleftarrow{C}_s = X) = d_p(n,k)$.

Remark 5.2. *Note that by Lemma 2.1 in* [12], *which is a special case of a theorem by McDiarmid,* [13], *our recursion for $d_p(k,n)$ also gives a formula for the probability that a given set of vertices X with $|X| = k$ is the connected component (or open cluster) containing s in $G(n, p/2)$.*

We now want to do something similar for the more complicated case of $g_n(p)$. For $n \geq 1$, $s \in X \subseteq [n]$, $s \in Y \subseteq [n]$ with $|X| = k, |Y| = m$ and $|[n] \setminus (Y \cup X)| = r$ define:

$$M_p(n,k,m,r) := \mathbb{P}_{\vec{G}(n,p)}(\overrightarrow{C}_s = X, \overleftarrow{C}_s = Y),$$

where in particular $M_p(1,1,1,0) = 1$.

Lemma 5.3. *We have the following recursions for M_p, where $k+m > n-r \geq k, m$ and $k, m \geq 1$*

(i) $M_p(n,k,m,r) = M_p(n-r,k,m,0)q^{r(r+k+m-n)}y^{r(2n-2r-k-m)}$, *for $r > 0$,*

(ii) $$M_p(n,k,m,0) = \sum_{j=1}^{n-k} \binom{n-k-1}{j-1} M_p(n-j,k,m-j,0) d_p(j,j) y^{j(n-m)} \cdot$$
$$\left(y^{m+k-n} - y^{n-k-j}q^{m+k-n}\right)q^{(j-1)(m+k-n)}y^{(j-1)(n-k-j)}, \quad \textit{for } n > k, n \geq m,$$

(iii) $M_p(n,k,m,r) = M_p(n,m,k,r)$,

(iv) $$M_p(n,n,n,0) = 1 - \sum_{j=1}^{n-1} \binom{n-1}{j-1} \sum_{k=j}^{n} \binom{n-j}{k-j} \sum_{m=j}^{n-k+j} \binom{n-k}{m-j} M_p(n,k,m,n-m-k+j),$$

(v) $$g_n(p) = \sum_{j=1}^{n-2} \sum_{k=j}^{n-1} \sum_{m=j}^{n-k+j} \binom{n-1}{j-1,\, m-j,\, k-j,\, n-k-m+j} \cdot$$
$$\frac{(n-k-1)(n-m-1)+n-j-1}{(n-1)(n-2)} M_p(n,k,m,n-m-k+j).$$

Proof. Assume, as given for the first equation, that $r > 0$. All vertices in $[n] \setminus (Y \cup X)$ must not have any edge directed to Y or from X. This means that there must be no edge at all to $Y \cap X$, which gives probability $q^{|[n]\setminus Y\cup X|\cdot|Y\cap X|} = q^{r(r+k+m-n)}$. There must not be any edge directed to $(Y \setminus X)$ and there must not be any edge directed from $(X \setminus Y)$. This gives a factor of $y^{|[n]\setminus Y\cup X|\cdot|(Y\setminus X)\cup(X\setminus Y)|} = y^{r(2n-2r-k-m)}$.

For equation (ii), note that $n > k$ and $r = 0$ imply that there exist a vertex $w \in Y \setminus X$. Let G be any directed graph on n vertices with $\overrightarrow{C}_s = X$ and $\overleftarrow{C}_s = Y$. If we remove vertex w and all its edges from G the resulting graph will still have $\overrightarrow{C}_s = X$ since $w \notin X$, whereas $\overleftarrow{C}_s = Y'$, for some Y' such that $Y \cap X \subseteq Y' \subseteq Y \setminus \{w\}$. Let $j = |Y \setminus Y'|$ and sum over all possible Y'. The probability is $M_p(n-j,k,m-j,0)$ that the subgraph on $[n] \setminus (Y \setminus Y')$ is as needed. The subgraph on $Y \setminus Y'$ must have $\overleftarrow{C}_w = Y \setminus Y'$ which gives probability $d_p(j,j)$. There must not be any edge directed from $X \setminus Y$ to $Y \setminus Y'$, since the vertices of the latter do not belong to X. The other direction is legal and this gives the factor $y^{(n-m)j}$. There must by the definition of Y' not be any edge at all between $X \cap Y$ and $Y \setminus (Y' \cup \{w\})$, which gives the factor $q^{(k+m-n)(j-1)}$. There must also not be any edge directed from $Y \setminus (Y' \cup \{w\})$ to $Y' \setminus (Y \cap X)$, which gives the factor $y^{(n-k-j)(j-1)}$. The only possible edges left to consider have one endpoint in w and the other in Y'. The edges between $Y' \setminus (Y \cap X)$ and w could have any direction and there must not be any edge directed from $Y \cap X$ to w (since $w \notin X$), but there must be at least one edge directed from w to Y' (since $w \in Y$). This gives probability $(y^{m+k-n} \cdot 1^{n-k-j} - q^{m+k-n} \cdot y^{n-k-j})$. Putting all this together gives formula (ii).

The third equation is obtained from the symmetry of reversing all directions.

The fourth equation follows from the fact that

$$\sum_{X,Y:s\in Y,X\subseteq[n]} \mathbb{P}_{\vec{G}(n,p)}(\overrightarrow{C}_s = X, \overleftarrow{C}_s = Y) = 1.$$

Here $j = |Y \cap X|$ and recall that $s \in Y \cap X$ is a necessary condition.

The last equation is obtained by summing over all possible pairs Y, X such that $a \notin Y, b \notin X$. Again $j = |Y \cap X|$ and the formula is split into the cases when $a \notin X$ and $a \in X$, respectively. We get

$$\begin{aligned} g_n(p) = \sum_{j=1}^{n-2} \binom{n-3}{j-1} \cdot \\ \Bigg(\sum_{k=j}^{n-2} \binom{n-2-j}{k-j} \sum_{m=j}^{n-k+j-1} \binom{n-k-1}{m-j} M_p(n,k,m,n-m-k+j) \\ + \sum_{k=j+1}^{n-1} \binom{n-2-j}{k-j-1} \sum_{m=j}^{n-k+j} \binom{n-k}{m-j} M_p(n,k,m,n-m-k+j) \Bigg), \end{aligned}$$

which after simplification gives the claimed formula. □

Note that in $\vec{G}(n,p)$ the functions $\mathbb{P}(a \not\to s)$ and $\mathbb{P}(a \not\to s, s \not\to b)$ are polynomials in p and hence continuous.

6. Computations and Conjectures for $\vec{G}(n,p)$

We have used `Maple` to compute the functions $f_n(p)$ and $g_n(p)$ for $n \leq 30$. Figure 1 displays the relative covariance $(g_n(p) - f_n(p)^2)/g_n(p)$. All curves start with being mildly negative. They then turn positive and for $n < 27$ they stay positive. For $n \geq 27$ however, they go below the p-axis again for some time.

For larger values of n it becomes infeasible for our computers to obtain the exact functions. We have instead for various fixed values of p used the recursions to obtain the value of the quotient $(g_n(p) - f_n(p)^2)/g_n(p)$ for $n \leq 300$. Based on these calculations we conjecture the following.

Conjecture 6.1. *For $n \geq 27$, the relative covariance changes sign at three critical probabilities $p_1(n) < p_2(n) < p_3(n)$.*

Conjecture 6.2. *Asymptotically, $p_1(n) \sim c_1/n$ and $p_2(n) \sim c_2/n$ for some constants c_1 and c_2.*

The computations indicate that very rough estimates of c_1 and c_2 are 0.36 and 7.5, respectively.

It follows from Remark 4.6 that there is a critical probability $p(n)$ such that $p(n) \to 1/2$ exponentially fast; if our conjectures hold, this critical probability is thus $p_3(n)$.

Conjecture 6.3. *For all $n \geq 27$, $p_3(n) < 1/2$.*

Conjecture 6.4. *For $n \geq 8$, the relative covariance $\frac{\mathbb{P}(a\not\to s, s\not\to b) - \mathbb{P}(a\not\to s)\,\mathbb{P}(s\not\to b)}{\mathbb{P}(a\not\to s, s\not\to b)} > \frac{2p-1}{3}$.*

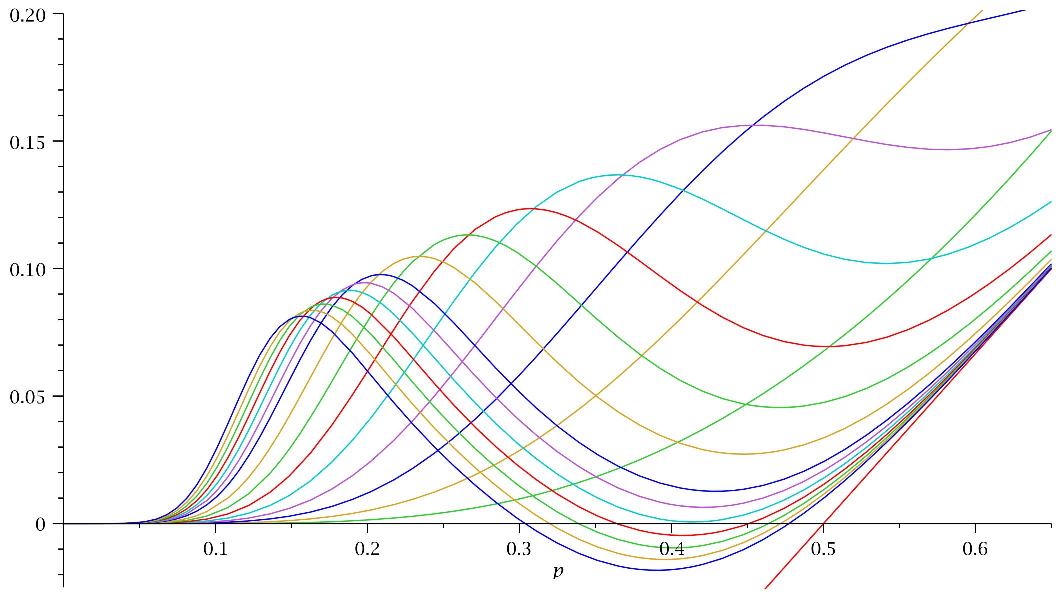

FIGURE 1. The relative correlation $\frac{\mathbb{P}(a\not\to s, s\not\to b)-\mathbb{P}(a\not\to s)\,\mathbb{P}(s\not\to b)}{\mathbb{P}(a\not\to s, s\not\to b)}$ in $\vec{G}(n,p)$ for $n = 8, 10, 12, 14, 16, 18, 20, 22, 24, 25, 26, 27, 28, 29, 30,$ and the asymptote $(2p-1)/3$. All curves are negative for very small p. For $n \geq 27$ we get three zeros.

Note that Conjecture 6.4 implies Conjecture 6.3 and the following related conjecture.

Conjecture 6.5. *For $p = \frac{1}{2}$, and $n \geq 6$ the relative covariance $\frac{\mathbb{P}(a\not\to s, s\not\to b)-\mathbb{P}(a\not\to s)\,\mathbb{P}(s\not\to b)}{\mathbb{P}(a\not\to s, s\not\to b)}$ is positive.*

7. Exact recursions in $\vec{G}(n,m)$

For convenience, let $A := \{a \not\to s\}$ and $B := \{s \not\to b\}$. In this section we will derive recursions for $\mathbb{P}(A)$ and $\mathbb{P}(A \cap B)$ in $\vec{G}(n,m)$ using the corresponding exact recursions in $\vec{G}(n,p)$ from Section 5.

Let as above $f_n(p) := \mathbb{P}_{\vec{G}(n,p)}(A) = \mathbb{P}_{\vec{G}(n,p)}(B)$ and $g_n(p) := \mathbb{P}_{\vec{G}(n,p)}(A \cap B)$, and let $h_n(m) := \mathbb{P}_{\vec{G}(n,m)}(A) = \mathbb{P}_{\vec{G}(n,m)}(B)$ and $k_n(m) := \mathbb{P}_{\vec{G}(n,m)}(A \cap B)$. Further, let $N := \binom{n}{2}$ be the number of edges in the complete graph K_n and $M \sim \mathrm{Bin}(N,p)$ be the actual number of edges in $G(n,p)$. Using that $G(n,m)$ can be seen as $\big(G(n,p) \mid M = m\big)$, we can express the functions $f_n(p)$ and $g_n(p)$ for $\vec{G}(n,p)$ using the corresponding functions, $h_n(m)$ and $k_n(m)$, for $\vec{G}(n,m)$.

$$f_n(p) = \sum_{m=0}^{N} \mathbb{P}(M = m) \cdot h_n(m) = \sum_{m=0}^{N} \binom{N}{m} p^m (1-p)^{N-m} \cdot h_n(m),$$
$$g_n(p) = \sum_{m=0}^{N} \mathbb{P}(M = m) \cdot k_n(m) = \sum_{m=0}^{N} \binom{N}{m} p^m (1-p)^{N-m} \cdot k_n(m).$$

These relations can be inverted by repeated differentiating.

Theorem 7.1.

$$h_n(m) = \frac{(N-m)!}{N!} f_n^{(m)}(0) - \sum_{i=0}^{m-1} \binom{m}{i} (-1)^{m-i} \cdot h_n(i),$$
$$k_n(m) = \frac{(N-m)!}{N!} g_n^{(m)}(0) - \sum_{i=0}^{m-1} \binom{m}{i} (-1)^{m-i} \cdot k_n(i).$$

Proof. It is sufficient to show one of the recursions.

$$\begin{aligned} f_n(p) &= \sum_{i=0}^{N} \binom{N}{i} p^i (1-p)^{N-i} \cdot h_n(i) \\ &= \sum_{i=0}^{N} \binom{N}{i} \sum_{j=0}^{N-i} \binom{N-i}{j} p^{i+j} (-1)^j \cdot h_n(i) \\ &= \sum_{k=0}^{N} \frac{N!}{(N-k)!} p^k \sum_{i=0}^{k} \frac{1}{i!(k-i)!} (-1)^{k-i} \cdot h_n(i). \end{aligned}$$

Differentiating m times and inserting $p = 0$ gives

$$f_n^{(m)}(0) = \frac{N!}{(N-m)!} m! \sum_{i=0}^{m} \frac{1}{i!(m-i)!} (-1)^{m-i} \cdot h_n(i)$$

from which we get

$$\sum_{i=0}^{m} \binom{m}{i} (-1)^{m-i} \cdot h_n(i) = \frac{(N-m)!}{N!} f_n^{(m)}(0),$$

which, after rearranging, gives the desired recursion. □

Computer calculations based on these recursions lead us to the following conjecture.

Conjecture 7.2. *For any fixed $n \geq 5$, the covariance in $\vec{G}(n, m)$ changes sign only once between two values of m.*

8. Comparison between $\vec{G}(n,p)$ and $\vec{G}(n,m)$

As before, let $A := \{a \not\to s\}$ and $B := \{s \not\to b\}$. For moderate n, the correlation between A and B is positive in $\vec{G}(n,p)$ for quite small p, while, for $\vec{G}(n,m)$ the proportion of links, m/N, where $N = \binom{n}{2}$, needs to be closer to 1 to get a positive correlation. In fact, we may study the conditional covariance given $M \sim \mathrm{Bin}(N,p)$ to show that the covariance in $\vec{G}(n,p)$ exceeds the average covariance in $\vec{G}(n,m)$ for fixed n and p.

Fact 8.1.

$$\mathrm{Cov}_{\vec{G}(n,p)}(A,B) = \mathbb{E}(\mathrm{Cov}(A,B \mid M)) + \mathrm{Var}(\mathbb{P}(A \mid M)) \tag{19}$$

where

$$\mathbb{E}(\mathrm{Cov}(A,B \mid M)) = \sum_{m=0}^{N} \mathbb{P}(M=m) \cdot \mathrm{Cov}_{\vec{G}(n,m)}(A,B).$$

To understand this statement (a standard type of variance analysis) we will as in Section 7 view $G(n,m)$ as $(G(n,p) \mid M = m)$. Note that

$$\begin{aligned}
\mathrm{Cov}_{\vec{G}(n,p)}(A,B) &= \mathbb{P}_{\vec{G}(n,p)}(A \cap B) - \mathbb{P}_{\vec{G}(n,p)}(A)\,\mathbb{P}_{\vec{G}(n,p)}(B) \\
&= \mathbb{E}(\mathbb{P}(A \cap B \mid M)) - \mathbb{E}(\mathbb{P}(A \mid M)\,\mathbb{P}(B \mid M)) \\
&\qquad + \mathbb{E}(\mathbb{P}(A \mid M)\,\mathbb{P}(B \mid M)) - \mathbb{E}(\mathbb{P}(A \mid M))\,\mathbb{E}(\mathbb{P}(B \mid M)) \\
&= \mathbb{E}(\mathrm{Cov}(A,B \mid M)) + \mathrm{Cov}(\mathbb{P}(A \mid M), \mathbb{P}(B \mid M))
\end{aligned}$$

and, as $\mathbb{P}(A \mid M) = \mathbb{P}(B \mid M)$, the formula follows.

The left-hand side of (19) is $\sim (2p-1)(1-p/2)^{2n-3}$ by Theorem 2.1. We can obtain the asymptotics of the two terms on the right-hand side too.

Theorem 8.2. *For every fixed $p \in (0,1]$,*

$$\mathbb{E}(\mathrm{Cov}(A,B \mid M)) = \Bigl(3 - (4-2p)e^{2\frac{p(1-p)}{(2-p)^2}} + o(1)\Bigr)\bigl(1-p/2\bigr)^{2n-3},$$

$$\mathrm{Var}(\mathbb{P}(A \mid M)) = 4\Bigl(e^{2\frac{p(1-p)}{(2-p)^2}} - 1 + o(1)\Bigr)\bigl(1-p/2\bigr)^{2n-2}.$$

Note that all three terms in (19), as well as $\mathbb{P}(A)\,\mathbb{P}(B) = \mathbb{P}(A)^2$, are of the same order (except when a term vanishes), viz. $(1-p/2)^{2n}$, see Theorem 2.1 and Lemma 4.2.

Proof. The sum of the right-hand sides equals $(2p-1+o(1))(1-p/2)^{2n-3}$, so by (1) and (19), it suffices to prove the second formula.

Let $\hat{p} = M/N = M/\binom{n}{2}$, and note that by the Law of Large Numbers, $\hat{p} \xrightarrow{\mathrm{p}} p$. We begin by observing that by the Central Limit Theorem,

$$n(\hat{p}-p) = \frac{n}{N^{1/2}} \cdot \frac{M-Np}{N^{1/2}} \xrightarrow{\mathrm{d}} \sqrt{2p(1-p)}Z,$$

where $\xrightarrow{\mathrm{d}}$ denotes convergence in distribution and $Z \sim N(0,1)$ is a standard normal variable. It follows that for any real constants a and b,

$$\frac{(1-\hat{p}/2)^{an+b}}{(1-p/2)^{an+b}} = \left(1 - \frac{\hat{p}-p}{2-p}\right)^{an+b} \xrightarrow{\mathrm{d}} \exp\left(a\frac{\sqrt{2p(1-p)}}{2-p}Z\right). \tag{20}$$

By Lemma 4.3,

$$\frac{\mathbb{P}(A \mid M)}{(1-\hat{p}/2)^{n-1}} = 2e^{-\frac{\hat{p}(1-\hat{p})}{(2-\hat{p})^2}} + o_{\mathrm{p}}(1) = 2e^{-\frac{p(1-p)}{(2-p)^2}} + o_{\mathrm{p}}(1),$$

and thus by (20)

$$\frac{\mathbb{P}(A \mid M)}{(1-p/2)^{n-1}} \xrightarrow{\mathrm{d}} 2e^{-\frac{p(1-p)}{(2-p)^2}} \exp\left(\frac{\sqrt{2p(1-p)}}{2-p}Z\right). \tag{21}$$

Denote the right-hand side of (21) by R. Since $\mathbb{E}\, e^{cZ} = e^{c^2/2}$ for any real c, we have

$$\mathbb{E}\, R^r = 2^r e^{(r^2-r)\frac{p(1-p)}{(2-p)^2}},$$

in particular, $\mathbb{E}\, R^2 = 4e^{2\frac{p(1-p)}{(2-p)^2}}$ and $\mathbb{E}\, R = 2$, so $\operatorname{Var} R = 4e^{2\frac{p(1-p)}{(2-p)^2}} - 4$. Hence the result follows if the variance converges in (21). For this, it suffices to show that

$$\mathbb{E}\left(\frac{\mathbb{P}(A \mid M)}{(1-p/2)^{n-1}}\right)^4 = O(1). \tag{22}$$

(See e.g. [7, Theorems 5.4.2 and 5.4.9] for this standard argument, and note that the same argument shows that all moments converge in (21).)

To verify (22), let $p_0 = p/2$. The proof of Lemma 4.3 shows that the estimates in Lemma 4.3 (with p replaced by $\hat{p}$) hold uniformly for $\hat{p} \geq p_0$. Since $q(n-1; n, m) \leq q'(n-1; n, \hat{p}) = (1-\hat{p}/2)^{n-1}$ by Lemma 3.2, this yields $\mathbb{P}(A \mid M) = O\big((1-\hat{p}/2)^{n-1}\big)$, provided $\hat{p} \geq p_0$. For $\hat{p} < p_0$, we simply use $\mathbb{P}(A \mid M) \leq 1$. Hence, for some constant C,

$$\mathbb{E}\left(\frac{\mathbb{P}(A \mid M)}{(1-p/2)^{n-1}}\right)^4 \leq (1-p/2)^{-4(n-1)}\, \mathbb{P}(\hat{p} < p_0) + C\, \mathbb{E}\left(\frac{1-\hat{p}/2}{1-p/2}\right)^{4(n-1)}. \tag{23}$$

The first term on the right-hand side is bounded since $\mathbb{P}(\hat{p} < p_0) = \mathbb{P}(M < Np_0) \leq e^{-Np/8}$ by the Chernoff bound [10, Theorem 2.1]. For the second term we have, letting $h = 8/(n(2-p))$ and recalling that M is a sum of N independent copies of $I \sim \operatorname{Bin}(1, p)$,

$$\begin{aligned}
\mathbb{E}\left(\frac{1-\hat{p}/2}{1-p/2}\right)^{4(n-1)} &= \mathbb{E}\left(1 - \frac{\hat{p}-p}{2-p}\right)^{4(n-1)} \leq \mathbb{E}\exp\left(-4(n-1)\frac{\hat{p}-p}{2-p}\right) \\
&= \mathbb{E}\exp\big(-h(M-Np)\big) = \big(\mathbb{E}\exp\big(-h(I-p)\big)\big)^N.
\end{aligned}$$

Now $\mathbb{E}\, e^{-h(I-p)} \leq e^{h^2/8}$, see [9, (4.16)] (a weaker estimate suffices), and thus the last expression is bounded by $e^{Nh^2/8} = O(1)$. This shows (22) and completes the proof. □

Numerical computations using the recursions of Sections 5 and 7 also suggest that all three quantities in (19), normalized by e.g. $(1-p/2)^{2n-3}$, converge quickly unless p is very small, cf. Remark 4.6; see Figure 2 for the case $n = 30$.

9. Quenched version

As mentioned in the introduction, we have so far studied the annealed model, i.e. the joint probability space of $\vec{G}(n,p)$ (or $\vec{G}(n,m)$) and that of the orientations. In the quenched model, the covariance is computed for each graph of $\vec{G}(n,p)$ (or $\vec{G}(n,m)$) and then averaged over all graphs.

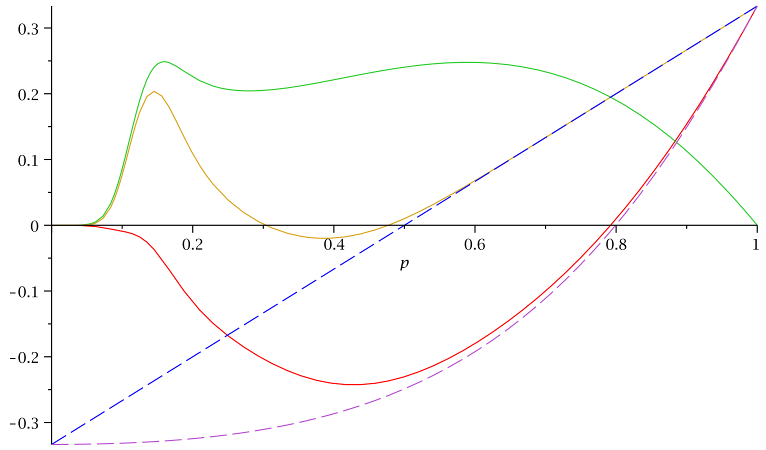

FIGURE 2. The solid curves are the covariance for $\vec{G}(30, p)$, middle curve, which by Fact 8.1 is the sum of the expected covariance for $\vec{G}(30, M)$, bottom, and $\mathrm{Var}(\mathbb{P}(A \mid M))$, top curve. The dashed curves are the asymptotic curves for $\vec{G}(n, p)$ (straight line) and expected covariance for $\vec{G}(n, M)$. All curves normalized with $3(1 - p/2)^{2n-3}$.

It is quite common that the results differ between the two models, and this seems to be the case here also. We have computed the quenched expectations for $\vec{G}(n, p)$ (and $\vec{G}(n, m)$) for small n ($n \leq 8$) and the covariances, as functions of p, look quite different, as can be seen from Figure 3 for $n = 8$.

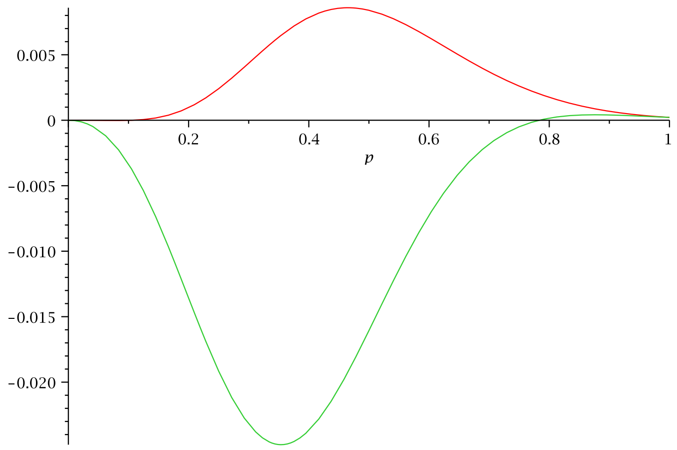

FIGURE 3. Annealed (top) and quenched (bottom) covariances for $\vec{G}(8, p)$.

For $n = 4, \dots, 8$ there is only one zero for the covariance, but this differs dramatically from the zero of the annealed model, as can be seen from Table 1.

n	Annealed	Quenched
4	1.000	1.000
5	0.729	0.927
6	0.276	0.857
7	0.152	0.809
8	0.107	0.783

TABLE 1. Zeroes of the covariances for $\vec{G}(n,p)$ in the annealed and quenched models.

Conditioning on the graph and taking expectations, we get that, similarly to Fact 8.1,

$$\mathrm{Cov}_{\vec{G}_a(n,p)}(A,B) = \mathrm{Cov}_{\vec{G}_q(n,p)}(A,B) + \mathrm{Cov}(\mathbb{P}(A \mid G), \mathbb{P}(B \mid G)), \tag{24}$$

where $\vec{G}_a(n,p)$ denotes the annealed model and $\vec{G}_q(n,p)$ denotes the quenched model; recall that $\mathrm{Cov}_{\vec{G}_q(n,p)}(A,B)$ is defined as $\mathbb{E}_{\vec{G}_q(n,p)}(\mathrm{Cov}(A,B \mid G))$. (A similar formula holds for $\vec{G}(n,m)$ and the corresponding quenched model.)

Here the conditional probabilities for A and B given the graph need not be equal, so that the last covariance in (24) could possibly be negative for some values of p and n. Even though our computations show that this is not the case when $n \leq 8$.

It is worth noting that, in contrast with the annealed model, in the quenched model $\vec{G}(n,p)$ and $\vec{G}(n,m)$ behave similarly, see Figure 4; this is not surprising, since the average taken in the quenched $\vec{G}(n,p)$ can be obtained by averaging over quenched $\vec{G}(n,m)$ with suitable weights. Also the quenched models appear, at least for small values of n, to be much closer to annealed $\vec{G}(n,m)$ than annealed $\vec{G}(n,p)$. In other words, the variation between different graphs with the same number of edges is of less importance than the variation caused by different number of edges. (The latter variation is quantified by Theorem 8.2.) This seems intuitively reasonable, since $\vec{G}(n,m)$ can be regarded as $\vec{G}(n,p)$ conditioned on the number of edges, which thus can be seen as a "semi-quenched" version.

Problem 9.1. *It would be interesting to find asymptotics for the quenched versions.*

References

[1] Sven Erick Alm & Svante Linusson, A counter-intuitive correlation in a random tournament, *Preprint 2009*, to appear in *Combinatorics, Probability and Computing*.
[2] Sven Erick Alm & Svante Linusson, Correlations for paths in random orientations of $G(n,p)$, *Preprint 2009* (earlier version of this paper), `arXiv:0906.0720`.
[3] Jörgen Backelin, Multinomial expressions summation asymptotic approximations, *Preprint 2010*.
[4] Jacob van den Berg, Olle Häggström & Jeff Kahn, Some conditional correlation inequalities for percolation and related processes, *Random Structures and Algorithms* **29** (2006), 417–435.
[5] Jacob van den Berg & Jeff Kahn, A correlation inequality for connection events in percolation, *Annals of Probability* **29** (2001), no. 1, 123–126.
[6] Geoffrey R. Grimmett, Infinite paths in randomly oriented lattices, *Random Structures and Algorithms* **18** (2001), no. 3, 257 – 266.
[7] A. Gut, *Probability: A Graduate Course*. Springer, New York, 2005.

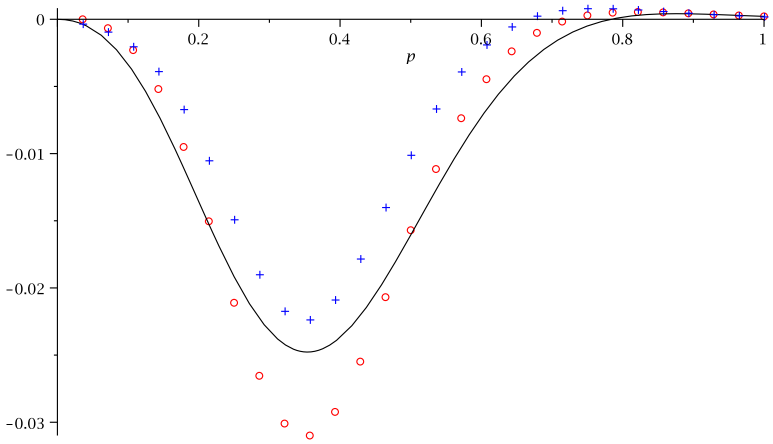

FIGURE 4. Covariances of A and B for $\vec{G}(8,p)$ (solid) and $\vec{G}(8,m)$ ($\circ$) in the quenched model and for $\vec{G}(8,m)$ ($+$) in the annealed model.

[8] Olle Häggström, Probability on bunkbed graphs, *Proceedings of FPSAC'03, Formal Power Series and Algebraic Combinatorics*, Linköping, Sweden, 2003. Available at `http://www.fpsac.org/FPSAC03/ARTICLES/42.pdf`

[9] W. Hoeffding, Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.* **58** (1963) 13–30.

[10] Svante Janson, Tomasz Łuczak & Andrzej Ruciński, *Random Graphs*, Wiley, New York, 2000.

[11] Svante Linusson, On percolation and the bunkbed conjecture, *Preprint 2008*, to appear in *Combinatorics, Probability and Computing*.

[12] Svante Linusson, A note on correlations in randomly oriented graphs, *Preprint 2009*, `arXiv:0905.2881`.

[13] Colin McDiarmid, General percolation and random graphs, *Adv. in Appl. Probab.* **13** (1981), 40–60.

Department of Mathematics, Uppsala University, P.O. Box 480, SE-751 06, Uppsala, Sweden.

E-mail address: `sea@math.uu.se`

Department of Mathematics, Uppsala University, P.O. Box 480, SE-751 06, Uppsala, Sweden.

E-mail address: `svante@math.uu.se`

Department of Mathematics, KTH-Royal Institute of Technology, SE-100 44, Stockholm, Sweden.

E-mail address: `linusson@math.kth.se`