

# On asymptotic properties of the rank of a special random adjacency matrix

Arup Bose

Indian Statistical Institute, Kolkata

Arnab Sen

University of California, Berkeley

## Abstract

Consider the matrix  $\Delta_n = ((I(X_i + X_j > 0)))_{i,j=1,2,\dots,n}$  where  $\{X_i\}$  are i.i.d. and their distribution is continuous and symmetric around 0. We show that the rank  $r_n$  of this matrix after proper centering and scaling behaves in some sense as a sum of a 1-dependent stationary sequence. As a consequence  $\sqrt{n}(r_n/n - 1/2)$  is asymptotically normal with mean zero and variance 1/4. We also show that  $n^{-1}r_n$  converges to 1/2 almost surely.

**Keywords.** Large dimensional random matrix, rank, almost sure representation, 1-dependent sequence, almost sure convergence, convergence in distribution.

**AMS 2000 Subject Classification.** Primary 60F99, Secondary 60F05, 60F15.

December 19, 2006

## 1. INTRODUCTION

Suppose  $\{X_1, X_2, \dots\}$  is a sequence of i.i.d. random variables. Define the symmetric matrix  $\Delta_n = ((I(X_i + X_j > 0)))_{i,j=1,2,\dots,n}$  where  $I$  is the indicator function.

The motivation for studying this matrix arises from the study of random network models, known as threshold models. Suppose that there is a collection of  $n$  nodes. Each node is assigned a fitness value and links are drawn among nodes when the total fitness crosses a threshold. This gives rise to a *good-get-richer* mechanism, in which sites with larger fitness are more likely to become hubs (i.e., to be connected). The scale free random network generated in this way is often used as a model in social networking, friendship networks, peer-to-peer (P2P) networks and networks of computer programs. Many features, such as power-law degree distributions, clustering, and short path lengths etc., of this random network has been studied in the physics literature extensively (see, for example, Caldarelli et. al. (2002), Söderberg (2002), Masuda et. al. (2005)).

Suppose that the fitness value of the sites are represented by the random variables  $X_i$  and we connect two points when their accumulated fitness is above the threshold 0. The matrix  $\Delta_n$  then represents the adjacency matrix of the above random graph.

Note that  $\Delta_n$  is a random matrix with zero and one entries. There are a few results known for the rank of random matrices with zero and one entries. See for example Costello and Yu (2006), Costello, Tao and Yu (2006). Here we give some interesting result on the rank of  $\Delta_n$ .

Suppose that the distribution of  $X_1$  is continuous and symmetric around 0. We then show that the rank  $r_n$  of this matrix is asymptotically half of the dimension. Indeed, the rank of  $\Delta_n$ , after proper centering can be approximated in distribution by the sum of a 1-dependent stationary sequence. As a consequence the rank is asymptotically normal with asymptotic mean  $n/2$  and variance  $n/4$ . This approximation is strong enough to also conclude that  $r_n/n$  converges to  $1/2$  almost surely.

**Theorem 1.** *Let  $\Delta_n$  be as above. Then, there exists a sequence of i.i.d.  $\text{Ber}(1, 1/2)$  random variables  $\xi_i$  such that with  $\Xi = ((\xi_{i \wedge j}))$ ,*

$$\text{rank}(\Delta_n) \stackrel{d}{=} \text{rank}(\Xi) \quad \text{and} \quad 0 \leq \text{rank}(\Xi) - 2 \sum_{i=1}^{n-1} I(\xi_i = 1, \xi_{i+1} = 0) \leq 1.$$

As a consequence,

- (A)  $0 \leq E(\frac{\text{rank}(\Delta_n)}{n} - 1/2) \leq \frac{1}{n}$
- (B)  $\sqrt{n}(\frac{\text{rank}(\Delta_n)}{n} - 1/2) \Rightarrow N(0, 1/4)$
- (C)  $\frac{\text{rank}(\Delta_n)}{n} \rightarrow 1/2$  almost surely.

To prove Theorem 1 we will need the following result which may be of independent interest.

**Lemma 1.** Suppose we have a (non-random) sequence  $\xi_1, \xi_2, \dots, \xi_n$  such that each  $\xi_i = 0$  or  $1$ . Let  $A = ((a_{ij}))_{1 \leq i, j \leq n}$  where  $a_{ij} = \xi_i \wedge j$ . Then

$$\text{rank}(A) - 2 \sum_{i=1}^{n-1} \mathbb{I}(\xi_i = 1, \xi_{i+1} = 0) = 0 \text{ or } 1.$$

**Proof:** The idea of the proof is as follows. First we apply an appropriate rank preserving transformation on  $A$  by permuting its rows and columns. Then we calculate the rank of the transformed matrix to get the result.

Suppose that  $k$  ( $0 \leq k \leq n$ ) many of  $\xi_i$ 's are non-zero. We can directly verify that the result holds for  $k = 0$  and  $k = n$ . Indeed, when  $k = 0$ ,  $\text{rank}(A) = 0$  and  $\sum_{i=1}^{n-1} \mathbb{I}(\xi_i = 1, \xi_{i+1} = 0) = 0$ . Similarly, if  $k = n$ ,  $\text{rank}(A) = 1$  and  $\sum_{i=1}^{n-1} \mathbb{I}(\xi_i = 1, \xi_{i+1} = 0) = 0$ .

**Claim:** Suppose that  $1 \leq k \leq n - 1$ . Let  $1 \leq m_1 < m_2 < \dots < m_k \leq n$  be such that

$$\xi_i = \begin{cases} 1 & \text{if } i \in \{m_1, m_2, \dots, m_k\}, \\ 0 & \text{otherwise.} \end{cases}$$

Then the matrix  $A$  may be reduced to the following matrix  $B = ((b_{ij}))$  by appropriate row column transformations.

$$b_{ij} = \begin{cases} 1 & \text{if } j \leq k, \quad i \geq m_j - j + 1 \\ 1 & \text{if } i \geq n - k + 1, \quad j \leq n - (m_{n-i+1} - n - i + 1) \\ 0 & \text{otherwise,} \end{cases}$$

i.e.,

$$B = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 \\ & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & \underbrace{1}_{m_1-1+1,1} & 0 & \vdots & 0 & 0 & 0 & 0 & \dots & 0 \\ & 1 & \underbrace{1}_{m_2-2+1,2} & \vdots & 0 & \dots & 0 & 0 & \dots & 0 \\ & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & 1 & 1 & \dots & 1 & \underbrace{1}_{n-1, n-m_2+2} & 0 & 0 & \dots & 0 \\ & 1 & 1 & \dots & 1 & 1 & \underbrace{1}_{n, n-m_1+1} & 0 & \dots & 0 \end{bmatrix}.$$

**Proof of the claim:** By definition of the matrix  $A$ , its  $m_1$ th row (resp. column) has all ones starting from the  $m_1$ th column (resp. row). Thus we may visualise  $A$  as follows:

$$A = \begin{bmatrix} 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & \underbrace{1}_{m_1, m_1} & 1 & \dots & 1 & 1 \\ 0 & \dots & 0 & 1 & a_{m_1+1, m_1+1} & \dots & a_{m_1+1, n-1} & a_{m_1+1, n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 1 & a_{n, m_1+1} & \dots & a_{n, n-1} & a_{n, n} \end{bmatrix}$$

We move the first  $(m_1 - 1)$  columns of  $A$  to the extreme east end and then move the  $m_1$ th row to the extreme south to get the following matrix  $A_1$  which has the same rank as that of the original matrix  $A$ .

$$A_1 = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ \underbrace{1}_{m_1, 1} & a_{m_1+1, m_1+1} & \dots & a_{m_1+1, n-1} & a_{m_1+1, n} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & a_{n, m_1+1} & \dots & a_{n, n-1} & a_{n, n} & 0 & \dots & 0 \\ 1 & 1 & \dots & 1 & \underbrace{1}_{1, n-m_1+1} & 0 & \dots & 0 \end{bmatrix}.$$

Now leave the first column, the last row, the first  $(m_1 - 1)$  zero rows and the last  $(m_1 - 1)$  zero columns intact and consider the remaining  $(n - m_1) \times (n - m_1)$  submatrix. This is a function of  $\xi_{m_1+1}, \xi_{m_1+2}, \dots, \xi_n$  and write it in the way that  $A$  was written and repeat the procedure. Note that now when we move the rows and columns, we move extreme south and east of only the submatrix but the remaining part of the matrix does not alter. It is easy to see that in  $(k - 1)$  more steps we obtain  $B$ . This proves the claim.

We return to the proof of the Lemma. Note that  $b_{ij} = 1$  iff  $b_{n-j+1, n-i+1} = 1$ . In other words,  $B$  is *anti-symmetric* (symmetric about its anti-diagonal). Indeed,  $B$  is the  $n \times n$  anti-symmetric 0-1 matrix containing minimum number of ones such that its  $i$ -th column contains  $n - m_i + i$  ones from bottom for  $1 \leq i \leq k$ .

Let  $u_i$  denote the number of 1's in the  $i$ -th column of  $B$ . Then,  $u_1 = n - m_1 + 1$ ,  $u_2 = n - m_2 + 2$ ,  $\dots$ ,  $u_k = n - m_k + k$ . Since  $m_1 < m_2 < \dots < m_k$  we have  $u_1 \geq u_2 \geq \dots \geq u_k$ . Also from the anti-symmetry of  $B$  we get for  $k < r \leq n$ ,  $u_r = \#\{i : 1 \leq i \leq k, u_i \geq r\}$ . This immediately implies that  $u_r \geq u_{r+1}$  for all  $r > k$ . Since  $u_k = n - m_k + k \geq k$  and  $u_{k+1} = \#\{i : 1 \leq i \leq k, u_i \geq k + 1\} \leq k$ , it follows that  $u_k \geq u_{k+1}$ . In fact  $u_k = u_{k+1}$  is not possible since then both are equal to  $k$  but then by definition of  $u_{k+1} < k$ . Thus  $\{u_i\}_{i=1}^n$  is a nonincreasing sequence.

Let  $d$  be the number of distinct elements of the set  $\{u_1, u_2, \dots, u_k\}$ . It is easy to see that  $d$  is equal to the number of occurrences of  $(1, 0)$  in  $\{(\xi_1, \xi_2), (\xi_2, \xi_3), \dots, (\xi_{n-1}, \xi_n), (\xi_n, 0)\}$ . Our goal is now to show that the rank of  $B$  is  $2d$  or  $2d - 1$ .

Let  $1 \leq i_1 < i_2 < \dots < i_d = k$  be such that  $u_1 = u_2 = \dots = u_{i_1} > u_{i_1+1} = \dots = u_{i_2} > u_{i_2+1} = \dots > \dots = u_k$ . We will consider the following two cases separately:

**Case I:**  $m_k < n$ . In this case  $u_k = (n - m_k) + k \geq k + 1$ . If we write out the whole  $u$ -sequence, it looks like this

$$\begin{array}{cccccccc} 1 \dots i_1 & i_1 + 1 \dots i_2 & \dots k & k + 1 \dots u_k & u_k + 1 \dots u_{i_{d-1}} & \dots u_{i_1} & u_{i_1} + 1 \dots n \\ u_{i_1} \dots u_{i_1} & u_{i_2} \dots u_{i_2} & \dots u_k & k \dots k & i_{d-1} \dots i_{d-1} & \dots i_1 & 0 \dots 0 \end{array}$$

The first row enumerates the column positions and the second row provides count of the number of one's in that column, that is, the corresponding  $u_i$ .

The rank of the matrix  $B$  can be easily computed by looking at the  $u$ -sequence. Each distinct non zero block contributes one to the rank. In this case,  $\text{rank}(B) = 2d$  and since  $\xi_n = 0$ , we have also  $d = \sum_{i=1}^{n-1} \mathbb{I}(\xi_i = 1, \xi_{i+1} = 0)$ . And therefore,

$$\text{rank}(B) = 2d = 2 \sum_{i=1}^{n-1} \mathbb{I}(\xi_i = 1, \xi_{i+1} = 0).$$

**Case II:**  $m_k = n$ . In this case  $u_k = (n - m_k) + k = k$ .

Now the  $u$ -sequence looks as follows

$$\begin{array}{cccccccc} 1 \dots i_1 & i_1 + 1 \dots i_2 & \dots k & k + 1 \dots u_k & u_k + 1 \dots u_{i_{d-1}} & \dots u_{i_1} & u_{i_1} + 1 \dots n \\ u_{i_1} \dots u_{i_1} & u_{i_2} \dots u_{i_2} & \dots u_k & i_{d-1} \dots i_{d-1} & i_{d-1} \dots i_{d-1} & \dots i_1 & 0 \dots 0 \end{array}$$

In this case,  $\text{rank}(B) = 2d - 1$  and since  $\xi_n = 1$ , we get  $\sum_{i=1}^{n-1} \mathbb{I}(\xi_i = 1, \xi_{i+1} = 0) = d - 1$ . So,

$$\text{rank}(B) - 2 \sum_{i=1}^{n-1} \mathbb{I}(\xi_i = 1, \xi_{i+1} = 0) = (2d - 1) - 2(d - 1) = 1.$$

This completes the proof of the Lemma.  $\square$

**Remark** Suppose that  $\xi_i$ 's are some 0-1 random variables. From the above proof it follows that  $Z = \text{rank}(B) - 2 \sum_{i=1}^{n-1} \mathbb{I}(\xi_i = 1, \xi_{i+1} = 0)$  is also a 0-1 valued random variable with  $P(Z = 1) = P(\xi_n = 1)$ .

**Proof of Theorem 1:** Write  $X_i$  as  $|X_i|S_i$  where  $S_i = \text{sgn}(X_i)$ . The assumption of continuous symmetric distribution of the  $\{X_i\}$  implies that  $\{|X_i|\}$  and  $\{S_i\}$  are independent. Let  $|X_{\sigma(1)}| < |X_{\sigma(2)}| < \dots < |X_{\sigma(n)}|$  be the ordered values of  $|X_i|$ ,  $i = 1, \dots, n$ .

Observe that the eigenvalues of  $\Delta_n$  are the same as the eigenvalues of  $((\mathbb{I}(X_{\sigma(i)} + X_{\sigma(j)} > 0)))$ . Note that  $\mathbb{I}(X_{\sigma(i)} + X_{\sigma(j)} > 0) = \mathbb{I}(S_{\sigma(i)\wedge j} = 1)$ . But since  $\sigma$  is a function of  $|X_i|$ ,  $i = 1, 2, \dots, n$  and  $\{S_i\}$ 's are independent of  $\{|X_i|\}$ 's, we have

$$\text{eigenvalues of } ((\mathbb{I}(S_{\sigma(i)\wedge j} = 1))) \stackrel{d}{=} \text{eigenvalues of } ((\mathbb{I}(S_{i\wedge j} = 1))).$$

Let us now define,

$$\xi_i = \mathbf{I}(S_i = 1), \quad i = 1, 2, \dots, n \text{ and } \Xi = ((\xi_i \wedge_j)).$$

Note that  $\xi_i \stackrel{i.i.d.}{\sim} \text{Ber}(1, 1/2)$ . From the discussion above, and from Lemma 1, we have

$$\text{rank}(\Delta_n) \stackrel{d}{=} \text{rank}(\Xi) \quad \text{and} \quad |\text{rank}(\Xi) - 2 \sum_{i=1}^{n-1} \mathbf{I}(\xi_i = 1, \xi_{i+1} = 0)| \leq 1.$$

Now note that  $Y_i = \{\mathbf{I}(\xi_i = 1, \xi_{i+1} = 0), i \geq 1\}$  is a stationary 1-dependent sequence of random variables with mean  $1/4$ . Part (A) now follows immediately.

From the discussion above, the asymptotic distribution of  $\sqrt{n}(\text{rank}(\Delta_n)/n - 1/2)$  is the same as that of  $n^{-1/2}(2 \sum_{i=1}^{n-1} \mathbf{I}(\xi_i = 1, \xi_{i+1} = 0) - n/2)$ . But by the central limit theorem for  $m$ -dependent stationary sequence, the latter is asymptotically normal with mean zero and variance  $\sigma^2$ . This variance is calculated as

$$\sigma^2 = (2)^2 [\text{Var}(Y_1) + 2 \text{Cov}(Y_1, Y_2)] = 4[(1/4 - 1/16) - 2/16] = 1/4.$$

This proves Part (B) of the theorem.

To prove Part (C), first observe that

$$\left| \frac{\text{rank}(\Xi)}{n} - \frac{2}{n} \sum_{i=1}^{n-1} \mathbf{I}(\xi_i = 1, \xi_{i+1} = 0) \right| \leq \frac{1}{n}.$$

By using the fact that the summands are all bounded, it is easy to check that the moment convergence in the central limit theorem for  $m$ -dependent sequences hold. That is

$$\mathbb{E} \left[ \sqrt{n} \left( \frac{2}{n} \sum_{i=1}^{n-1} \mathbf{I}(\xi_i = 1, \xi_{i+1} = 0) - 1/2 \right) \right]^4 \rightarrow 3/16.$$

Combining all the above, we conclude that

$$\mathbb{E} \left[ \sqrt{n} \left( \frac{\text{rank}(\Delta_n)}{n} - \frac{1}{2} \right) \right]^4 \leq \mathbb{E} \left[ \sqrt{n} \left( \frac{2}{n} \sum_{i=1}^{n-1} \mathbf{I}(\xi_i = 1, \xi_{i+1} = 0) - \frac{1}{2} \right) \right]^4 + O(1) = O(1).$$

This implies that

$$\sum_{n=1}^{\infty} \mathbb{E} \left( \text{rank}(\Delta_n)/n - 1/2 \right)^4 < \infty$$

and thus now Part (C) follows from the Borel Cantelli Lemma.  $\square$

**Acknowledgement** We are grateful to Anish Sarkar for helpful discussions and comments.

#### REFERENCES

- [1] Caldarelli, G., Capocci, A., Rios, P. De Los, and Muñoz, M.A. (2002). Scale-free networks from varying vertex intrinsic fitness. *Physical Review Letters*, 89, 258702.
- [2] Costello, Kevin, Tao, Terence and Vu, Van. Random symmetric matrices are almost surely non-singular. Preprint.
- [3] Costello, Kevin and Vu, Van. (2006). The rank of random graphs. arXiv:math:PR/0606414 v1.
- [4] Masuda, Naoki, Miwa, Hiroyoshi, and Konno, Norio (2005). Geographical threshold graphs with small-world and scale-free properties *Physical Review E*, 71, 036108.
- [5] Söderberg, Bo (2002). General formalism for inhomogeneous random graphs *Physical Review E*, 66 066121.

Arup Bose  
Indian Statistical Institute, Kolkata  
203 B. T. Road  
Kolkata 700108  
INDIA  
abose@isical.ac.in

Arnab Sen  
Department of Statistics  
University of California, Berkeley  
CA 94720  
USA  
arnab@stat.berkeley.edu