Elect. Comm. in Probab. **12** (2007), 200–205

ELECTRONIC COMMUNICATIONS in PROBABILITY

ON ASYMPTOTIC PROPERTIES OF THE RANK OF A SPECIAL RANDOM ADJACENCY MATRIX

ARUP BOSE
Statistics and Mathematics Unit, Indian Statistical Institute, 202 B.T. Road, Kolkata 700108, India
email: abose@isical.ac.in

ARNAB SEN
Department of Statistics, University of California, Berkeley, CA 94720, USA
email: arnab@stat.berkeley.edu

Submitted December 19, 2006, *accepted in final form* May 1, 2007

AMS 2000 Subject classification: Primary 60F99, Secondary 60F05, 60F15
Keywords: Large dimensional random matrix, rank, almost sure representation, 1-dependent sequence, almost sure convergence, convergence in distribution.

Abstract

Consider the matrix $\Delta_n = ((\ \mathrm{I}(X_i + X_j > 0)\ ))_{i,j=1,2,\ldots,n}$ 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 is equal in distribution to $2\sum_{i=1}^{n-1} \mathrm{I}(\xi_i = 1, \xi_{i+1} = 0) + \mathrm{I}(\xi_n = 1)$ where $\xi_i \overset{i.i.d.}{\sim} \mathrm{Ber}(1, 1/2)$. 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.

1 Introduction

Suppose $\{X_1, X_2, \ldots\}$ is a sequence of i.i.d. random variables. Define the symmetric matrix $\Delta_n = ((\ \mathrm{I}(X_i + X_j > 0)\ ))_{i,j=1,2,\ldots,n}$ 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 Δ_n then represents the adjacency matrix of the above random graph.

Note that Δ_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 Vu (2006), Costello, Tao and Vu (2006). In particular, the latter shows that the matrix with i.i.d. entries which are Bernoulli with probability 1/2 each is almost surely nonsingular as $n \to \infty$.

Suppose that the distribution of X_1 is continuous and symmetric around 0. Then the entries of the matrix Δ_n are identically distributed, being Bernoulli with probability 1/2 but now there is strong dependency among them. We show that the rank r_n of this matrix is asymptotically half of the dimension. Indeed, the rank of Δ_n, can be approximated in distribution by the sum of a 1-dependent stationary sequence. As a consequence it is asymptotically normal with asymptotic mean $n/2$ and variance $n/4$. Further, r_n/n converges to 1/2 almost surely. In what follows, $X \stackrel{\mathcal{D}}{=} Y$ means the two random variables X and Y have the same probability distributions.

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

$$\mathrm{rank}(\Delta_n) \stackrel{\mathcal{D}}{=} \mathrm{rank}(\Xi) \; = 2\sum_{i=1}^{n-1} \mathrm{I}(\xi_i = 1, \xi_{i+1} = 0) + \mathrm{I}(\xi_n = 1).$$

As a consequence,

(A) $\mathrm{E}\left(\frac{\mathrm{rank}(\Delta_n)}{n}\right) = 1/2$

(B) $\sqrt{n}\big(\frac{\mathrm{rank}(\Delta_n)}{n} - 1/2\big) \Rightarrow N(0, 1/4)$

(C) $\frac{\mathrm{rank}(\Delta_n)}{n} \to 1/2$ *almost surely.*

Proof of Theorem 1: Let

$$S_i = \begin{cases} +1 & \text{if } X_i \geq 0 \\ -1 & \text{if } X_i < 0 \end{cases}.$$

The assumption of continuous symmetric distribution of the $\{X_i\}$ implies that the two sequences $\{|X_i|\}$ and $\{S_i\}$ are independent. Let $(\sigma(1), \ldots, \sigma(n))$ be the random permutation of $(1, \ldots, n)$ such that $|X_{\sigma(1)}| < |X_{\sigma(2)}| < \cdots < |X_{\sigma(n)}|$ are the ordered values of $|X_i|, \;\; i = 1, 2, \ldots, n$.

Since any row column permutation leaves the eigenvalues unchanged, the eigenvalues of Δ_n are the same as the eigenvalues of the matrix $((\mathrm{I}(X_{\sigma(i)} + X_{\sigma(j)} > 0)))$. Note that $\mathrm{I}(X_{\sigma(i)} + X_{\sigma(j)} > 0) = \mathrm{I}(S_{\sigma(i\wedge j)} = 1)$. But since σ is a function of $|X_i|, \; i = 1, 2, \ldots n$ and $\{S_i\}$'s are independent of $\{|X_i|\}$'s, we have the following equality in distribution:.

$$\text{eigenvalues of } ((\; \mathrm{I}(S_{\sigma(i\wedge j)} = 1) \;)) \stackrel{\mathcal{D}}{=} \text{eigenvalues of } ((\; \mathrm{I}(S_{i\wedge j} = 1) \;)).$$

Let us now define,

$$\xi_i = \mathrm{I}(S_i = 1), \; i = 1, 2, \ldots, n \text{ and } \; \Xi = ((\xi_{i\wedge j})).$$

Note that $\xi_i \stackrel{i.i.d.}{\sim} \mathrm{Ber}(1, 1/2)$.

To complete the proof, we will need the following Lemma We will first finish the proof of the theorem assuming the lemma and then prove the lemma.

Lemma 1. *Suppose we have any sequence* $\xi_1, \xi_2, \ldots, \xi_n$ *such that each* $\xi_i = 0$ *or* 1. *Let* $\mathbf{A} = ((a_{ij}))_{1 \le i,j \le n}$ *where* $a_{ij} = \xi_{i \wedge j}$. *Then*

$$\operatorname{rank}(\mathbf{A}) - 2\sum_{i=1}^{n-1} \mathrm{I}(\xi_i = 1, \xi_{i+1} = 0) = \mathrm{I}(\xi_n = 1).$$

From the discussion above, and from Lemma 1, we have

$$\operatorname{rank}(\Delta_n) \stackrel{\mathcal{D}}{=} \operatorname{rank}(\Xi) \ = 2\sum_{i=1}^{n-1} \mathrm{I}(\xi_i = 1, \xi_{i+1} = 0) + \mathrm{I}(\xi_n = 1) \tag{1}$$

where $\xi_i \stackrel{i.i.d.}{\sim} \mathrm{Ber}(1, 1/2)$.

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

From equation (1) above, the asymptotic distribution of $\sqrt{n}\big(\operatorname{rank}(\Delta_n)/n - 1/2\big)$ is the same as that of $n^{-1/2}\big(2\sum_{i=1}^{n-1} \mathrm{I}(\xi_i = 1, \xi_{i+1} = 0) - n/2\big)$. But by the central limit theorem for m-dependent stationary sequence (see for example Brockwell and Davis, 1990, page 213), the latter is asymptotically normal with mean zero and variance σ^2. This variance is calculated as

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

This proves Part (B) of the theorem.

To prove Part (C), first observe that

$$\Big|\frac{\operatorname{rank}(\Xi)}{n} - \frac{2}{n}\sum_{i=1}^{n-1} \mathrm{I}(\xi_i = 1, \xi_{i+1} = 0)\Big| \le \frac{1}{n}.$$

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

$$\mathrm{E}\Big[\sqrt{n}\Big(\frac{2}{n}\sum_{i=1}^{n-1} \mathrm{I}(\xi_i = 1, \xi_{i+1} = 0) - 1/2\Big)\Big]^4 \to 3/16.$$

Combining all the above, we conclude that

$$\mathrm{E}\Big[\sqrt{n}\Big(\frac{\operatorname{rank}(\Delta_n)}{n} - \frac{1}{2}\Big)\Big]^4 \le \mathrm{E}\Big[\sqrt{n}\Big(\frac{2}{n}\sum_{i=1}^{n-1} \mathrm{I}(\xi_i = 1, \xi_{i+1} = 0) - \frac{1}{2}\Big)\Big]^4 + O(1) = O(1).$$

This implies that

$$\sum_{n=1}^{\infty} \mathrm{E}\Big(\operatorname{rank}(\Delta_n)/n - 1/2\Big)^4 < \infty$$

and now Part (C) follows from Borel Cantelli Lemma, proving the theorem completely. □

Proof of Lemma 1: The idea of the proof is as follows. First we apply an appropriate rank preserving transformation on $\mathbf{A}$ by permuting its rows and columns. Then we calculate the rank of the transformed matrix to get the result.

Suppose that k $(0 \le k \le n)$ many of ξ_i's are non-zero. If $k = 0$, then $\text{rank}(A) = 0$, $\xi_n = 0$ and $\sum_{i=1}^{n-1} \mathrm{I}(\xi_i = 1, \xi_{i+1} = 0) = 0$. Similarly, if $k = n$, then $\text{rank}(A) = 1$, $\xi_n = 1$ and $\sum_{i=1}^{n-1} \mathrm{I}(\xi_i = 1, \xi_{i+1} = 0) = 0$. Hence the result holds for $k = 0$ and $k = n$.

Claim: Suppose that $1 \le k \le n-1$. Let $1 \le m_1 < m_2 < \cdots < m_k \le n$ be such that

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

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

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

i.e.,

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

Proof of the claim: By definition of the matrix $\mathbf{A}$, its m_1th row (resp. column) has all ones starting from the m_1th column (resp. row). Thus we may visualise $\mathbf{A}$ as follows:

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

We move the first $(m_1 - 1)$ columns of A to the extreme east end and then move the m_1th row to the extreme south to get the following matrix A_1 which has the same rank as that of the original matrix $\mathbf{A}$.

$$\mathbf{A}_1 = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ \underbrace{1}_{m_1,1} & a_{m_1+1,m_1+1} & \dots & a_{m_1+1,n-1} & a_{m_1+1,n} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & a_{n,m_1+1} & \dots & a_{n,n-1} & a_{n,n} & 0 & \dots & 0 \\ 1 & 1 & \dots & 1 & \underbrace{1}_{1,n-m_1+1} & 0 & \dots & 0 \end{bmatrix}.$$

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 $\mathbf{A}$ 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 $\mathbf{B}$. 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, $\mathbf{B}$ is *anti-symmetric* (symmetric about its anti-diagonal). Indeed, $\mathbf{B}$ 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 $1\le i\le k$.

Let u_i denote the number of 1's in the i-th column of $\mathbf{B}$. 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 $u_1\ge u_2\ge\dots\ge u_k$. Also from the anti-symmetry of $\mathbf{B}$ we get for $k<r\le n$, $u_r = \#\{i : 1\le i\le k,\ \ u_i\ge r\}$. This immediately implies that $u_r\ge u_{r+1}$ for all $r>k$. Since $u_k = n-m_k+k\ge k$ and $u_{k+1} = \#\{i : 1\le i\le k,\ \ u_i\ge k+1\}\le k$, it follows that $u_k\ge u_{k+1}$. 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 $\{u_1,u_2,\cdots,u_k\}$. 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 $\mathbf{B}$ is $2d$ or $2d-1$.

Let $1\le i_1<i_2<\dots<i_d=k$ 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 $u_k=(n-m_k)+k\ge k+1$. If we write out the whole u-sequence, it looks like this

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

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

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

Case II: $m_k = n$**.** In this case $u_k = (n - m_k) + k = k$. Now the u-sequence looks as follows

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

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

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

This completes the proof of the Lemma. □

Acknowledgement. We are grateful to Anish Sarkar for helpful discussions and comments. We are also grateful to the Referees who have sent detailed comments, leading to a much improved presentation. Finally, we thank the Editor for a very quick report.

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. (2006) Random symmetric matrices are almost surely non-singular. *Duke Math J.* Volume 135, Number 2, 395-413 MR2267289

[3] Costello, Kevin and Vu, Van. (2006). The rank of random graphs. *arXiv:math:PR/0606414 v1.*

[4] Brockwell, Peter J. and Davis, Richard A. (1991). *Time series: theory and methods*, Second edition. *Springer Series in Statistics. Springer Verlag, New York.* MR1093459

[5] Masuda, Naoki, Miwa, Hiroyoshi, and Konno, Norio (2005). Geographical threshold graphs with small-world and scale-free properties *Physical Review E*, 71, 036108.

[6] Söderberg, Bo (2002). General formalism for inhomogeneous random graphs *Physical Review E*, 66 066121. MR1953933