Acta Math. Hungar., **134 (1–2)** (2012), 45–53
DOI: 10.1007/s10474-011-0125-z
First published online June 17, 2011

WHEN THE DEGREE SEQUENCE IS A SUFFICIENT STATISTIC

V. CSISZÁR[1,*,†], P. HUSSAMI[2], J. KOMLÓS[3,§], T. F. MÓRI[1,*,‡], L. REJTŐ[2,4,¶] and G. TUSNÁDY[2,†]

[1]Eötvös Loránd University, Budapest, Hungary
e-mails: villo@ludens.elte.hu, moritamas@ludens.elte.hu

[2]Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences, Budapest, Hungary
e-mails: huprim@yahoo.com, tusnady@renyi.hu

[3]Rutgers University, Department of Mathematics, New Brunswick, New Jersey, USA
e-mail: komlos@math.rutgers.edu

[4]University of Delaware, Statistics Program, FREC, CANR, Newark, Delaware, USA
e-mail: rejto@udel.edu

(Received October 21, 2010; revised January 24, 2011; accepted February 15, 2011)

Abstract. There is a uniquely defined random graph model with independent adjacencies in which the degree sequence is a sufficient statistic. The model was recently discovered independently by several authors. Here we join to the statistical investigation of the model, proving that if the degree sequence is in the interior of the polytope defined by the Erdős–Gallai conditions, then a unique maximum likelihood estimate exists.

1. Introduction

In a version of Albert–Barabási random graphs [10] the adjacencies are independent with probabilities

$$p_{t,u} = \frac{\kappa}{\sqrt{tu}},$$

where κ is a positive parameter. Observing that the right hand side is a product, the generalization is straightforward with a slight modification. We

* Supported by the European Union and the European Social Fund under the grant agreement no. TÁMOP 4.2.1./B-09/KMR-2010-0003.
† Supported by OTKA grant K 76481.
‡ Supported by OTKA grant K 67961.
§ Supported in part by NSF grant DMS-0902241.
¶ Corresponding author.

Key words and phrases: degree sequence of graphs, random graph, sufficient statistics, maximum likelihood estimation.
2000 Mathematics Subject Classification: primary 62F10, secondary 05C07.

supplant the probabilities with odds:

$$\frac{p_{t,u}}{1-p_{t,u}} = \alpha_t \alpha_u, \tag{1}$$

where $\alpha_1, \dots, \alpha_T$ are arbitrary positive parameters assigned to the vertices of the graph. The probability of a graph G is thus

$$P(G;\alpha) = C(\alpha) \prod_{t=1}^{T} \alpha_t^{d_t}, \tag{2}$$

where d_t denotes the degree of vertex t and

$$C(\alpha) = \prod_{t=2}^{T} \prod_{u=1}^{t-1} \frac{1}{1+\alpha_t \alpha_u}.$$

$P(G;\alpha)$ depends on the observed graph G through the degree sequence only, thus the degree sequence is a sufficient statistic.

A simple random model like this ought to have a history. We refer to [3] where the origin and basic properties of the model are delineated. There the parametrization $\beta_t = \log \alpha_t$ is used and the random graph is called *β-model,* what we also adapt here. The β-model is investigated also in [1] and [2].

The structure of the paper is the following. In the next section, we prove our main results regarding maximum likelihood estimation in the β-model. Theorem 2.1 is very close to Theorem 1.4 and Theorem 2.9 is very close to Theorem 1.5 of [3]. However, our proofs appear to be more natural and simpler. The novelty of our Theorem 2.9 is that there is no need for *a priori* assuming the existence of the ML estimate, since in Theorem 2.8 we settle a conjecture of Diaconis et al. [3] about when the ML estimate exists. An extended discussion of our results is given in [8]. Statistical inference on discrete structures is developed in [6] and for random permutations in [5].

2. Maximum likelihood estimation in the β-model

Throughout the paper, $T \geqq 3$ will be a fixed integer, and we will deal with simple undirected graphs on the vertex set $\{1,\dots,T\}$. The edge set of a graph G is denoted by $E(G)$. A *random graph* will formally stand for an arbitrary probability distribution on the set of all $2^{\binom{T}{2}}$ graphs.

It is well-known (see [7] and [11]) that a sequence $\{d_t,\ 1 \leqq t \leqq T\}$ of nonnegative integers is the degree sequence of a graph if and only if $\sum_{t=1}^{T} d_t$

is even, and the following Erdős–Gallai conditions hold:

$$\sum_{t=1}^{k} d_{\pi(t)} \leqq k(k-1) + \sum_{u=k+1}^{T} \min\left(k, d_{\pi(u)}\right), \quad k = 1, \dots, T-1 \tag{3}$$

for any permutation $\pi(1), \dots, \pi(T)$ of integers $1, \dots, T$. How about the expected degree sequence of a random graph? To this end, let $\varepsilon_{t,u}$ denote the indicator of the edge between vertices u and t. Of course, $\varepsilon_{t,u} = \varepsilon_{u,t}$ and $\varepsilon_{t,t} = 0$. Moreover, let $p_{t,u} = P(\varepsilon_{t,u} = 1)$ be the probability of an edge between u and t. Then the expected degree of vertex t is $\Delta_t = \sum_{u=1}^{T} p_{t,u}$. If we write both $p_{t,u}$ and Δ_t as vectors, we obtain the concise notation $\Delta = Sp$, where $S : \mathbb{R}^{\binom{T}{2}} \to \mathbb{R}^T$ is a linear transformation.

THEOREM 2.1. *There exists a random graph with vertex set* $\{1, \dots, T\}$ *such that the expected degree of vertex* t *is* Δ_t, *if and only if* $\Delta_t \geqq 0$ *and*

$$\sum_{t=1}^{k} \Delta_{\pi(t)} \leqq k(k-1) + \sum_{u=k+1}^{T} \min\left(k, \Delta_{\pi(u)}\right), \quad k = 1, \dots, T-1 \tag{4}$$

hold true for any permutation $\pi(1), \dots, \pi(T)$ *of integers* $1, \dots, T$.

PROOF. In one direction, let k be a fixed integer as in (4) and define

$$\sigma_s = \sum_{t=1}^{k} p_{\pi(t),\pi(s)}, \qquad 1 \leqq s \leqq T.$$

The left hand side of (4) equals $\sum_{s=1}^{T} \sigma_s$. For $s \leqq k$ we have that $\sigma_s \leqq k-1$ and for $s > k$ we have $\sigma_s \leqq k$ and $\sigma_s \leqq \Delta_{\pi(s)}$. Thus Δ_t satisfies (4).

In the other direction, let us denote the set of T-dimensional vectors with nonnegative coordinates satisfying (4) by D_T. The expected degree sequence Δ of a random graph is the convex combination of the degree sequences satisfying the Erdős–Gallai conditions (3). Let us denote this set by R_T. The conditions (3) can be rewritten as a set of linear inequalities, namely

$$\sum_{t=1}^{k} d_{\pi(t)} \leqq k(k-1) + \sum_{u=k+1}^{T} \left(\lambda_u k + (1-\lambda_u) d_{\pi(u)}\right)$$

must hold for all k, all permutations π, and all sequences $(\lambda_{k+1}, \dots, \lambda_T) \in \{0,1\}^{T-k}$. This shows that the conditions remain valid after taking convex combinations, thus $R_T \subseteqq D_T$. We want to determine all maximal faces of R_T. These are determined by inequalities $a^T \Delta \leqq b_a$, where $a \in \mathbb{R}^T$ is a suitable vector, and b_a a corresponding constant. For any a, we get for the supporting hyperplane

$$a^T\Delta = \sum_{t=1}^{T} a_t\Delta_t = \sum_{u<t} p_{tu}(a_t + a_u) \leqq \sum_{a_t+a_u>0} (a_t + a_u) = b_a.$$

Clearly, b_a can be attained, and is attained by exactly the points $\Delta = Sp$ with

(5) $\quad p_{tu} = 1$ if $a_t + a_u > 0$, $p_{tu} = 0$ if $a_t + a_u < 0$, $p_{tu} \in [0, 1]$ otherwise.

We claim that the maximal faces of R_T have normal vectors with $a_t \in \{-1, 0, +1\}$. Indeed, for any vector a, let $c = \operatorname{sgn}(a)$. Observe that if Δ is on the supporting hyperplane with normal vector a, i.e. $a^T\Delta = b_a$, then in view of (5), it is also on the supporting hyperplane with normal vector c, i.e. $c^T\Delta = b_c$. This proves our claim. Moreover, the inequalities $a^T\Delta \leqq b_a$, with $a_t \in \{-1, 0, 1\}$ are evidently equivalent with the Erdős–Gallai conditions (4), finishing the proof. □

REMARK 2.2. Theorem 2.1 is not new. First it was published by M. Koren (see Theorem 1 in [9]). Our proof is more straightforward and simpler.

In the β-model, according to (2), the conditional distribution of the graph, given that its degree sequence is $d = \{d_1, \ldots, d_T\}$, is the uniform distribution on the set of all graphs with this degree sequence, for any degree sequence d. The converse also holds.

THEOREM 2.3. *Suppose that for a random graph with independent adjacencies* $0 < r_{tu} = \frac{p_{tu}}{1-p_{tu}} < \infty$. *If, for any degree sequence d, the conditional distribution of the graph, given that its degree sequence is d, is the uniform distribution, then the random graph belongs to the β-model.*

PROOF. For $T \leqq 3$, the theorem trivially holds. If $T \geqq 4$, let $d = (2, \ldots, 2)$, this is an interior point of R_T. Suppose that the distribution of a random graph, conditioned on its degree sequence being d, is uniform. This implies that for all graphs G with degree sequence d,

$$\prod_{(tu)\in E(G)} r_{tu} = \kappa$$

for some constant $\kappa > 0$. It is easily seen that for any four distinct vertices u, t, v, w, there exists a realization G of $d = (2, \ldots, 2)$ such that $(ut), (vw) \in E(G)$ but $(uv), (tw) \notin E(G)$. Then we can make a swap showing that $r_{ut}r_{vw} = r_{uv}r_{tw}$. This in turn implies that the definition

$$\alpha_u = \sqrt{\frac{r_{ut}r_{uv}}{r_{tv}}}$$

does not depend on the choice of t, v, and with this definition $r_{ut} = \alpha_u \alpha_t$. $\square$

REMARK 2.4. We conjecture that Theorem 2.3 holds for any individual degree sequence in the interior of D_T.

Now we turn to maximum likelihood estimation in the β-model. Suppose we observe the graph G, and we want to maximize the likelihood $P(G; \alpha)$. We will show in several steps that if the degree sequence of G lies in the interior of D_T, then a unique maximum likelihood estimate exists, and we give a simple iteration converging to this maximum likelihood estimate.

To start out, the likelihood equations are given by

$$\frac{\partial}{\partial \alpha_t} \log P(G; \alpha) = \frac{d_t}{\alpha_t} - \sum_{u \neq t} \frac{\alpha_u}{1 + \alpha_u \alpha_t} = 0, \quad 1 \leqq t \leqq T.$$

THEOREM 2.5. *Let $\Delta = \{\Delta_1, \ldots, \Delta_T\}$ be an interior point of D_T. Then, the likelihood equations*

$$\sum_{u \neq t} \frac{\alpha_t \alpha_u}{1 + \alpha_t \alpha_u} = \Delta_t, \quad 1 \leqq t \leqq T \tag{6}$$

have a solution.

PROOF. In the β-model, the lefthand side of (6) is just $\sum_{u \neq t} p_{t,u}$, i.e. the expected degree of vertex t. We shall maximize the entropy

$$H(p) = - \sum_{1 \leqq u < t \leqq T} \bigl(p_{t,u} \log p_{t,u} + (1 - p_{t,u}) \log (1 - p_{t,u}) \bigr) \tag{7}$$

of an arbitrary random graph with independent adjacencies, fixing the expected degrees of the vertices, in other words we require

$$\sum_{u \neq t} p_{t,u} = \Delta_t, \qquad 1 \leqq t \leqq T. \tag{8}$$

Since Δ lies in the interior of D_T, there exists a solution of (8) such that $0 < p_{t,u} < 1$ holds true for all $1 \leqq u < t \leqq T$. On the other hand on the boundary of the cube $C = [0,1]^{\binom{T}{2}}$, at least one of the probabilities $p_{t,u}$ is equal to either 0 or 1, where the (one-sided) partial derivative $\frac{\partial}{\partial p_{t,u}} H(p)$ of the entropy equals $\pm\infty$ while the corresponding term $-\bigl(p_{t,u} \log p_{t,u} + (1 - p_{t,u}) \log (1 - p_{t,u})\bigr)$ turns from a positive number to zero. Thus the maximum of $H(p)$, given (8) is in the interior of D_T.

Let us denote the Lagrangian multipliers with β_t and set

$$\tilde{H}(p) = H(p) + \sum_{t=1}^{T} \beta_t \Big(\sum_{u \neq t} p_{t,u} - \Delta_t \Big).$$

At the maximum place, the partial derivatives

$$\frac{\partial \tilde{H}(p)}{\partial p_{t,u}} = -\log \frac{p_{t,u}}{1 - p_{t,u}} + \beta_t + \beta_u$$

should be zero. It means that $p_{t,u}$ has the form (1). □

The likelihood equations are

$$\alpha_t = d_t \Big(\sum_{u \neq t} \frac{1}{\alpha_u^{-1} + \alpha_t} \Big)^{-1}, \quad 1 \leqq t \leqq T.$$

Let $\mathbb{R}_+^T$ be the T dimensional space with positive coordinates. For any α, let $\varphi : \mathbb{R}_+^T \to \mathbb{R}_+^T$ be defined by

$$\varphi_t(\alpha) = d_t \Big(\sum_{u \neq t} \frac{1}{\alpha_u^{-1} + \alpha_t} \Big)^{-1} \qquad 1 \leqq t \leqq T. \tag{9}$$

The solutions of the likelihood equation are the fixed points of the map φ. Starting from any $\alpha^{(0)}$, we can run the iteration $\alpha^{(n+1)} = \varphi\big(\alpha^{(n)}\big)$, hoping to converge to the maximum likelihood estimate.

THEOREM 2.6. *For any $x, y \in \mathbb{R}_+^T$ define*

$$\varrho(x,y) = \max \left(\max_{1 \leqq t \leqq T} \frac{x_t}{y_t}, \ \max_{1 \leqq t \leqq T} \frac{y_t}{x_t} \right). \tag{10}$$

Then for $x \neq y$

$$\varrho\Big(\varphi\big(\varphi(x)\big), \varphi\big(\varphi(y)\big)\Big) < \varrho(x,y). \tag{11}$$

We shall need the following lemma.

LEMMA 2.7. *For any integer $n > 1$ and arbitrary positive numbers $a_1, \dots, a_n$ and $b_1, \dots, b_n$ we have*

$$\frac{a_1 + \dots + a_n}{b_1 + \ldots + b_n} \leqq \max_{1 \leqq i \leqq n} \frac{a_i}{b_i}. \tag{12}$$

Equality holds true if and only if the ratios $\frac{a_i}{b_i}$ have the same value.

PROOF. Let $\kappa > 0$ be such that $a_i \leqq \kappa b_i$ hold true for $i = 1, \dots, n$. Then the left hand side of (12) can not be larger than κ and the only way to equality is $a_i = \kappa b_i$ for all i. ☐

PROOF OF THEOREM 2.6. Let us denote $\varrho(x, y)$ by κ. Let us fix t and apply Lemma 2.7 for $n = T$, $a_i = \left(y_i^{-1} + y_t\right)^{-1}$, $b_i = \left(x_i^{-1} + x_t\right)^{-1}$:

$$\frac{\varphi_t(x)}{\varphi_t(y)} = \frac{\sum_{u \neq t} \left(y_u^{-1} + y_t\right)^{-1}}{\sum_{u \neq t} \left(x_u^{-1} + x_t\right)^{-1}} \leqq \max_{u \neq t} \frac{x_u^{-1} + x_t}{y_u^{-1} + y_t}. \tag{13}$$

Applying Lemma 2.7 again we get that

$$\max_{u \neq t} \frac{x_u^{-1} + x_t}{y_u^{-1} + y_t} \leqq \max_{u \neq t} \max\left(\frac{y_u}{x_u}, \frac{x_t}{y_t}\right) \leqq \kappa. \tag{14}$$

Interchanging x and y, we see that $\varrho\big(\varphi(x), \varphi(y)\big) \leqq \varrho(x, y)$. Equality in (14) can hold only if $\frac{x_t}{y_t} = \kappa$ and $\frac{y_u}{x_u} = \kappa$ at least for one $u \neq t$. But then the equality in (13) can be valid only if $\frac{y_u}{x_u} = \kappa$ for all $u \neq t$. Thus a necessary condition for $\varrho\big(\varphi\big(\varphi(x)\big), \varphi\big(\varphi(y)\big)\big) = \kappa$ would be that $\max \frac{\varphi_t(x)}{\varphi_t(y)} = \kappa$ and $\min \frac{\varphi_t(x)}{\varphi_t(y)} = 1/\kappa$. But this cannot happen, because the first one implies $\frac{y_t}{x_t} = \kappa$ for all but one t, while the second implies $\frac{y_t}{x_t} = 1/\kappa$ for all but one t. ☐

THEOREM 2.8. *If the degree sequence of the graph G lies in the interior of D_T, then there exists a unique parameter vector $(\alpha_1, \dots, \alpha_T)$ satisfying* (6), *which is also the unique maximizer of the likelihood function* (2).

PROOF. If the degree sequence is in the interior of D_T then we first show that there is indeed at least one maximum of the likelihood function

$$P(G; \alpha) = P(G; p) = \prod_{(t,u) \in E(G)} p_{t,u} \prod_{(t,u) \notin E(G)} (1 - p_{t,u}),$$

where p is calculated from α via (1). Indeed, by Theorem 2.5 we can parametrize the β-model with the expected degree sequences $\Delta = Sp$. Suppose now that Δ approaches the boundary of D_T. By continuity, it suffices to show that if p^* is such that $\Delta^* = Sp^*$ is on the boundary of D_T, then $P(G; p^*) = 0$. Denote the degree sequence of G by d^*. Since Δ^* is on the boundary, there are real numbers $(x_0, x_1, \dots, x_T)$ such that

$$x_0 + \sum_{t=1}^{T} x_t d_t \geqq 0$$

for all degree sequences $d \in D_T$, and

$$E_{p^*}\left(x_0 + \sum_{t=1}^{T} x_t d_t\right) = x_0 + \sum_{t=1}^{T} x_t \Delta_t^* = 0.$$

Since

$$x_0 + \sum_{t=1}^{T} x_t d_t^* > 0,$$

$P(G; p^*)$ must be zero.

Since the maxima satisfy (6), Theorem 2.6 ensures the uniqueness. $\square$

THEOREM 2.9. *Let φ be defined by* (9). *If d lies in the interior of D_T, then the iteration $\alpha^{(n+1)} = \varphi\big(\alpha^{(n)}\big)$ starting with arbitrary $\alpha^{(0)} \in \mathbb{R}_+^T$ converges at a geometric rate, and*

$$\alpha_t = \lim_{n\to\infty} \alpha_t^{(n)}, \quad t = 1, \dots, T$$

is the unique maximum likelihood estimate.

PROOF. We showed in Theorem 2.6 that φ is a contraction in the metric $r(x, y) = \log \varrho(x, y)$. Starting from any $\alpha^{(0)}$, the sequence of iterates remains bounded. Moreover, any limit point of the sequence is a fixed point of φ. But due to the contractive property, the only fixed point is the maximum likelihood estimate guaranteed by Theorem 2.8. The geometric rate of convergence follows from a standard compactness argument. $\square$

We remark at this point that the β-model can naturally be studied in the framework of exponential families. The random graphs

$$\mathcal{E} = \left\{ P_\beta : \, P_\beta(G) = C(\beta) \exp\left(\sum_{t=1}^{T} \beta_t d_t(G) \right) \right\}$$

form an exponential family. For any $\Delta \in D_T$, the corresponding linear family is given by

$$\mathcal{L} = \big\{ Q : \, E_Q\big(d_t(G)\big) = \Delta_t, \; 1 \leqq t \leqq T \big\}.$$

We can now use Section 3 of [4], in particular Theorems 3.2 and 3.3. These ensure that the intersection $\mathcal{L} \cap \mathrm{cl}\,(\mathcal{E})$ consists of a single random graph P^*, and if there exists a random graph Q in $\mathcal{L}$ with full support, then $P^* \in \mathcal{E}$. If Δ is an interior point of D_T, then indeed there exists an interior point p of the unit cube with $Sp = \Delta$, which ensures that the corresponding random graph has full support. Thus we obtain another proof of Theorem 2.5 (and the uniqueness of the solution is also proved). By the Pythagorean theorem of information geometry, P^* is also the unique maximum likelihood estimate, thus proving Theorem 2.8 again.

Acknowledgement. We are thankful to the referee informing us about Koren's paper.

References

[1] A. Barvinok and J. A. Hartigan, An asymptotic formula for the number of non-negative integer matrices with prescribed row and column sums, preprint (2009), http://arxiv.org/abs/0910.2477.

[2] A. Barvinok and J. A. Hartigan, The number of graphs and a random graph with a given degree sequence, preprint (2010), http://arxiv.org/abs/1003.0356.

[3] S. Chatterjee, P. Diaconis and A. Sly, Random graphs with a given degree sequence, preprint (2010, arXiv: 1005.1136v3 [math.PR]).

[4] I. Csiszár and P. Shields, Information theory and statistics: A tutorial, *Foundations and Trends in Communications and Information Theory*, **1**(4) (2004), 417–528. now Publishers

[5] V. Csiszár, Conditional independence relations and log-linear models for random matchings, *Acta Math. Hungar.*, **122** (2009), 131–152.

[6] V. Csiszár, L. Rejtő and G. Tusnády, Statistical inference on random structures, in: E. Győri et al. (Eds.), *Horizon of Combinatorics* (2008), pp. 37–67.

[7] P. Erdős and T. Gallai, Graphs with given degrees of vertices, *Mat. Lapok*, **11** (1960), 264–274 (in Hungarian).

[8] P. Hussami, Statistical inference on random graphs, PhD Thesis, 2010 (submitted to Central European University, Budapest).

[9] M. Koren, Extreme degree sequences of simple graphs, *J. Combinatorial Theory B*, **15** (1973), 213–224.

[10] M. Newman, A.-L. Barabási and D. Watts, *The Structure and Dynamics of Networks*, Princeton Studies in Complexity, Princeton University Press (2007).

[11] G. Sierskma and H. Hoogeveen, Seven criteria for integer sequences being graphic, *J. Graph Theory*, **2** (1991), 223–231.