Asymptotic degree distributions in large (homogeneous) networks: A little theory and a counterexample$^{ab}$

Armand M. Makowski

University of Maryland at College Park (and National Science Foundation (CISE/CCF/CIF))

---

$^a$Supported by NSF Grant CCF-1217997 and a Lady Davis Trust Fellowship.

$^b$Joint work with Siddarth Pal (ECE, now with Raytheon BBN/ARL)
References


Good generative random graph models

In many surveys we are told: At the beginning there were Erdős-Rényi graphs …

In Erdős-Rényi graphs, the degree distribution is “Poisson”-like

Story line?
Erdős-Rényi graphs $\mathbb{G}(n; p)$:

$$D_{n,k}(p) =_{st} \text{Bin}(n - 1; p), \quad n = 2, 3, \ldots$$

$$0 \leq p \leq 1$$

Homogeneity – There is a (common) degree distribution:

$$D_n(p) =_{st} D_{n,k}(p), \quad k = 1, 2, \ldots, n$$

Theoretical justification via Poisson convergence – For some $\lambda > 0$,

$$D_n(p_n) \xrightarrow{\text{n}} \mathbf{p}_\lambda \quad \text{if} \ p_n \sim \frac{\lambda}{n}$$

where $\mathbf{p}_\lambda$ is the Poisson pmf on $\mathbb{N}$ with parameter $\lambda$. 
We are also told

Erdős-Rényi graph models are deemed inappropriate for many “complex networks” because observations/measurements point to a degree distribution which is not Poisson-like.

(Vague empirical statement) For large networks (large $n$),

$$\frac{N_n(d)}{n} \sim C d^{-\alpha} \quad (1)$$

for some $\alpha$ in the range $[2, 3]$ and $C > 0$ with

$$N_n(d) = \text{Number of nodes with degree } d \text{ in } G_n$$

Power/Pareto-like degree distribution. Scale-free networks.
Preferential attachment models  
(Barabási and Albert (1999))

Generative model based on growth and preferential attachment – Markov sequence of random graphs

\[ G_n = (V_n, E_n), \quad n = 1, 2, \ldots \]

No more a single node degree distribution but instead an (empirical) networkwide degree distribution:

\[ P_n(d) = \frac{N_n(d)}{n}, \quad n = 1, 2, \ldots \]

\[ d = 0, 1, \ldots \]
with

\[ N_n(d) = \text{Number of nodes with degree } d \text{ in } \mathbb{G}_n \]

It can be shown that

\[ P_n(d) \xrightarrow{P} p_{\text{SF}}(d) \quad d = 0, 1, \ldots \]

with pmf \( p_{\text{SF}}(d) = (p_{\text{SF}}(d), \ d = 0, 1, \ldots) \) on \( \mathbb{N} \) such that

\[ p_{\text{SF}}(d) \sim Cd^{-3} \quad (d \to \infty) \]

Original model

Barabási and Albert (1999), Bollobás and Riordan (2000) – Many variations on this theme
Back to Erdős-Rényi graphs

Apples vs. oranges?

With scaling $p : \mathbb{N}_0 \to [0, 1]$, identically distributed rvs

$$D_{n,k}(p_n), \quad k = 1, 2, \ldots, n$$

with a common distribution $\text{Bin}(n - 1, p_n)$ vs. an empirical measure

$$P_n(d) = \frac{1}{n} \sum_{k=1}^{n} 1[D_{n,k}(p_n) = d], \quad d = 0, 1, \ldots$$

Is there a relationship (in the large $n$ limit) between the pmf of the generic degree and the empirical degree pmf?
Resolving the difference

Lemma 1 If $p_n \sim \frac{\lambda}{n}$ for some $\lambda > 0$, then for each $d = 0, 1, \ldots$,

$$\frac{1}{n} \sum_{k=1}^{n} 1[D_{n,k}(p_n) = d] \xrightarrow{P} p_\lambda(d)$$

where $p_\lambda = (p_\lambda(d), d = 0, 1, \ldots)$ is the Poisson pmf on $\mathbb{N}$ with parameter $\lambda$ given by

$$p_\lambda(d) = \frac{\lambda^d}{d!} e^{-\lambda}, \quad d = 0, 1, \ldots$$
Why should one care? To be able to establish a **meaningful** comparison between Erdős-Rényi graph models and preferential attachment models.

---

**A natural question:** Can we generalize Lemma 1 to a larger class of random graphs models?
A general framework
(and an easy result)
Consider a sequence of random graphs \( \{G_n, n = 2, 3, \ldots \} \), in the **homogeneous** case, with

\[
D_{n,1} =_{st} D_{n,2} =_{st} \ldots =_{st} D_{n,n} =_{st} D_n.
\]

**Q:** If \( D_n \xrightarrow{n} D \), i.e.,

\[
\lim_{n \to \infty} \mathbb{P} [D_n = d] = \mathbb{P} [D = d], \quad d = 0, 1, \ldots
\]

then is it true that

\[
\frac{N_n(d)}{n} \xrightarrow{P} n \mathbb{P} [D = d], \quad d = 0, 1, \ldots
\]

where

\[
N_n(d) = \sum_{k=1}^{n} 1 [D_{n,k} = d]?
\]
A: Yes provided

- Homogeneity
- Existence of an asymptotic degree distribution
- Asymptotic uncorrelatedness

These conditions hold for many classes of *generative* random graph models
Assumption 1 (Homogeneity) For each \( n = 2, 3, \ldots \), the degree rvs in \( \mathbb{G}_n \) are equidistributed in the sense that

\[
D_{n,k} =_{st} D_{n,1}, \quad k = 1, 2, \ldots, n
\]

and

\[
(D_{n,k}, D_{n,\ell}) =_{st} (D_{n,1}, D_{n,2}) \quad k \neq \ell \quad k, \ell = 1, \ldots, n
\]

Assumption 2 (Existence of an asymptotic degree distribution)  
Under Assumption 1, there exists an \( \mathbb{N} \)-valued rv \( D \) such that

\[
D_{n,1} \xrightarrow{\text{d}} D.
\]

Let \( \mathbf{p} = (p(d), \ d = 0, 1, \ldots) \) denote the pmf of the limiting rv \( D \).
Assumption 3 (Asymptotic uncorrelatedness) Under Assumption 1, for each \( d = 0, 1, \ldots \), the rvs \( 1[D_{n,1} = d] \) and \( 1[D_{n,2} = d] \) are asymptotically uncorrelated in the sense that

\[
\lim_{n \to \infty} \text{Cov} [1[D_{n,1} = d], 1[D_{n,2} = d]] = 0.
\]

Assumptions 1 and 2 are the baseline assumptions.

Assumption 3 amounts to

\[
\lim_{n \to \infty} (\mathbb{P}[D_{n,1} = d, D_{n,2} = d] - \mathbb{P}[D_{n,1} = d] \mathbb{P}[D_{n,2} = d]) = 0.
\]
Main (but easy) result

Proposition 1 Under Assumptions 1-3, we have

$$\frac{N_n(d)}{n} \overset{p}{\rightarrow}_n p(d), \quad d = 0, 1, \ldots$$

where the pmf $\mathbf{p} = (p(d), \; d = 0, 1, \ldots)$ is postulated in Assumption 2.
Here, there is **equivalence** between $L^2$-convergence and convergence in probability:

- Convergence in probability implies $L^2$-convergence by bounded convergence
- $L^2$ convergence implies convergence in probability by Tchebychev’s inequality: With $\varepsilon > 0$, Tchebychev’s inequality gives

$$
\Pr \left[ \left| \frac{N_n(d)}{n} - p(d) \right| > \varepsilon \right] \leq \frac{1}{\varepsilon^2} \cdot \mathbb{E} \left[ \left| \frac{N_n(d)}{n} - p(d) \right|^2 \right].
$$

Thus, as we **mimic** the proof of WLLNs, we need only prove that

$$
\lim_{n \to \infty} \mathbb{E} \left[ \left| \frac{N_n(d)}{n} - p(d) \right|^2 \right] = 0, \quad d = 0, 1, \ldots
$$
Lemma 2 If Assumptions 1-2 hold, then

\[
\lim_{n \to \infty} \mathbb{E} \left[ \left| \frac{N_n(d)}{n} - p(d) \right|^2 \right] = \lim_{n \to \infty} \text{Cov} \left[ \mathbf{1}[D_n, 1 = d], \mathbf{1}[D_n, 2 = d] \right]
\]

for each \( d = 0, 1, \ldots \) with the understanding that if one of the limits exists, so does the other and the limiting values coincide.

Fix \( n = 2, 3, \ldots \) and \( d = 0, 1, \ldots \): Centering about the mean, we have

\[
\mathbb{E} \left[ \left| \frac{N_n(d)}{n} - p(d) \right|^2 \right] = \text{Var} \left[ \frac{N_n(d)}{n} \right] + \mathbb{E} \left[ \left| \frac{N_n(d)}{n} \right| - p(d) \right]^2.
\]
\[ \mathbb{E} [N_n(d)] = \mathbb{E} \left[ \sum_{k=1}^{n} 1[D_{n,k} = d] \right] \]
\[ = \sum_{k=1}^{n} \mathbb{P} [D_{n,k} = d] = n \cdot \mathbb{P} [D_{n,1} = d] \]

Assumption 1

so that

\[ \lim_{n \to \infty} \mathbb{E} \left[ \frac{N_n(d)}{n} \right] = \lim_{n \to \infty} \mathbb{P} [D_{n,1} = d] = p(d) \]

Assumption 2

\[ \lim_{n \to \infty} \left| \mathbb{E} \left[ \frac{N_n(d)}{n} \right] - p(d) \right|^2 = 0 \]
Var \[ N_n(d) \] = n \cdot \text{Var} \[ 1[D_{n,1} = d] \] 
+ n(n - 1) \cdot \text{Cov} \[ 1[D_{n,1} = d] , 1[D_{n,2} = d] \]

Assumption 1

so that

\[
\text{Var}\left[ \frac{N_n(d)}{n} \right] = \frac{1}{n} \cdot \text{Var} \[ 1[D_{n,1} = d] \] 
+ \frac{n - 1}{n} \cdot \text{Cov} \[ 1[D_{n,1} = d], 1[D_{n,2} = d] \]
\]

\[
\lim_{n \to \infty} n^{-1} \text{Var} \[ 1[D_{n,1} = d] \] = 0
\]

Always
Combining, under Assumption 1 and Assumption 2, we conclude that

\[
\lim_{n \to \infty} \mathbb{E} \left[ \left| \frac{N_n(d)}{n} - p(d) \right|^2 \right] \\
= \lim_{n \to \infty} \left( \text{Var} \left[ \frac{N_n(d)}{n} \right] + \mathbb{E} \left[ \frac{N_n(d)}{n} \right]^2 - p(d) \right) \\
= \lim_{n \to \infty} \text{Var} \left[ \frac{N_n(d)}{n} \right] \\
= \lim_{n \to \infty} \text{Cov} \left[ \mathbf{1} \left[ D_{n,1} = d \right], \mathbf{1} \left[ D_{n,2} = d \right] \right]
\]
The desired result

$$\lim_{n \to \infty} \mathbb{E} \left[ \left| \frac{N_n(d)}{n} - p(d) \right|^2 \right] = 0$$

is **equivalent** to Assumption 3

---

**Corollary 1** Under Assumptions 1-2, we have

$$\frac{N_n(d)}{n} \xrightarrow{p_n} p(d), \quad d = 0, 1, \ldots$$

where the pmf $p = (p(d), \ d = 0, 1, \ldots)$ is postulated in Assumption 2 **if and only if** Assumptions 3 holds, namely

$$\lim_{n \to \infty} \text{Cov} \left[ \mathbf{1} \left[ D_{n,1} = d \right], \mathbf{1} \left[ D_{n,2} = d \right] \right] = 0$$
Convergence of
\[ \left\{ \frac{N_n(d)}{n}, \ n = 1, 2, \ldots \right\} \]
in either mode **fails** if
\[ \lim_{n \to \infty} \text{Cov} \left[ 1[D_{n,1} = d], 1[D_{n,2} = d] \right] > 0 \]
A commonly occurring framework
Given is an underlying **parametric** family of random graphs

\[ \{ \mathbb{G}(n; \alpha), \ n = 2, 3, \ldots \}, \quad \alpha \in A \subset \mathbb{R}^p \]

Typically, for each \( \alpha \in A \), the degree rvs \( D_{n,1}(\alpha), \ldots, D_{n,n}(\alpha) \) are **exchangeable**.

There **exists** a scaling \( \alpha^* : \mathbb{N}_0 \to A \) such that

\[ D_{n,1}(\alpha^*_n) \implies D \]

for some **non**-degenerate \( \mathbb{N} \)-valued rv – Existence of a **maximal** component
Set

\[ G_n = G(n; \alpha^*_n), \quad n = 1, 2, \ldots \]

Assumptions 1 and 2 are automatically satisfied, so that only Assumption 3 need to be verified

The result applies to a long list of models:

- Erdős-Rényi graphs \( G(n; p) \)
  
  \[ p_n^* \sim \frac{\lambda}{n} \]

- Geometric random graphs \( G(n; \rho) \)
  
  \[ \pi (\rho_n^*)^2 \sim \frac{\lambda}{n} \]

- Random key graphs \( K(n; \theta) \)
  
  \[ \frac{(K_n^*)^2}{P_n^*} \sim \frac{\lambda}{n} \]

- \( \ldots \)

Always?
Random threshold graphs
(A counterexample)
Motivation

Preferential attachment model is based on growth and preferential attachment

- Cumulative advantage – “The rich get richer”
- Predicated on information about the degree of each vertex being available to newly added nodes, either explicitly or implicitly – Questionable assumption in some cases

An answer to this modeling issue: Hidden variable models (Caldarelli et al. (2002))

- Creation of a link between two nodes expresses a mutual benefit based on intrinsic attributes
Random threshold graphs

Fitness – I.i.d. rvs \( \{\xi, \xi_i, \ i = 1, 2, \ldots, n\} \)

“The rich know the richer” – With threshold \( \theta \) in \( \mathbb{R} \),

\[
\mathbb{T}(n; \theta) : \ i \sim j \ \text{if and only if} \ \xi_i + \xi_j > \theta
\]

Non-negative fitness – The fitness rvs \( \xi_1, \ldots, \xi_n \) are assumed to be non-negative i.i.d. rvs — We need consider only \( \theta > 0 \).
Already a large body of work concerning this class of random graph models, e.g., degree distribution, clustering, degree correlations, etc.

In particular, power law for the degree distribution can emerge in the many node limit \((n \to \infty)\) when \(\xi \sim \text{Exp}(\lambda)\) under scaling \(\theta^* : \mathbb{N}_0 \to \mathbb{R}_+\) such that

\[
\theta^*_n = \lambda^{-1} \log n, \quad n = 1, 2, \ldots
\]
Fujihara et al. (2003) show that $D_{n,1}(\theta_n^*) \xrightarrow{n} D$, namely

$$
\lim_{n \to \infty} \mathbb{P}[D_{n,1}(\theta_n^*) = d] = p_{Fuj}(d), \quad d = 0, 1, \ldots
$$

with

$$
p_{Fuj}(d) = \mathbb{E} \left[ \frac{(e^{\lambda \xi})^d}{d!} e^{-e^{\lambda \xi}} \right], \quad d = 0, 1, \ldots
$$

**Conditional Poisson pmf**

(with rate $e^{\lambda \xi}$)

---

**Claim (Caldarelli et al.):** Random threshold graphs with exponential fitness provide an **alternative** to the preferential attachment model due to the fact that

$$
p_{Fuj}(d) \sim d^{-2} \quad (d \to \infty)
$$

Beware!
Applying the theory ($\xi \sim \text{Exp}(\lambda)$)

Here we focus on

$$\mathbb{G}_n = \mathbb{T}(n; \theta^*_n), \quad n = 1, 2, \ldots$$

Assumption 1 holds: Trivial by exchangeability

Assumption 2 holds with $p = p_{\text{Fuj}}$ as shown by Fujihara et al. (2003)

Assumption 3 fails!
Formally,

**Proposition 2** With $\xi \sim \text{Exp}(\lambda)$ for some $\lambda > 0$, for each $d = 0, 1, \ldots$, the limit

$$C(d) \equiv \lim_{n \to \infty} \text{Cov}[1[D_{n,1}(\theta^*_n) = d], 1[D_{n,2}(\theta^*_n) = d]]$$

exists with $C(d) > 0$.

For instance,

$$C(0) = \mathbb{E}\left[e^{-\max(e^{\lambda \xi_1}, e^{\lambda \xi_2})}\right] - \mathbb{E}\left[e^{-(e^{\lambda \xi_1} + e^{\lambda \xi_2})}\right] > 0$$

since $\max(e^{\lambda \xi_1}, e^{\lambda \xi_2}) < e^{\lambda \xi_1} + e^{\lambda \xi_2}$
Figure 1: Empirical degree distribution for individual runs, empirical distribution averaged over runs, and theoretical nodal distribution.
Figure 2:
Limiting empirical distribution?

Theorem 1  For each $d = 0, 1, \ldots$, 

$$
\frac{N_n(d)}{n} \xrightarrow{n} \Pi(d)
$$

where $\Pi(d)$ is a non-degenerate $[0, 1]$-valued rv with

$$
\mathbb{E}[\Pi(d)] = p_{\text{Fuj}}(d) \quad \text{and} \quad \text{Var}[\Pi(d)] > 0.
$$
Basic ideas of the proof

- Multi-dimensional version of Fujihara et al. – For each $r = 1, 2, \ldots$,

$$\lim_{n \to \infty} \mathbb{P} [D_{n,1} = d_1, \ldots, D_{n,r} = d_r]$$

- Method of moments:

$$\lim_{n \to \infty} \mathbb{E} \left[ \left( \frac{N_n(d)}{n} \right)^T \right] = \lim_{n \to \infty} \mathbb{P} [D_{n,1} = d, \ldots, D_{n,r}(d)]$$

- Characteristic function

$$\lim_{n \to \infty} \mathbb{E} \left[ e^{it \frac{N_n(d)}{n}} \right], \quad t \in \mathbb{R}$$
Non-homogeneous random networks

...
Consider the sequence of random graphs \( \{G_n, \ n = 1, 2, \ldots\} \) 
without any additional assumptions.

For each \( n = 2, 3, \ldots \), write \( V_n = \{1, \ldots, n\} \) and set

\[
\Sigma_n \equiv \{(k, \ell) \in V_n \times V_n : k \neq \ell\}
\]

Consider the rv \((\nu_n, \mu_n) : \Omega \to \Sigma_n\) which is uniformly distributed over \(\Sigma_n\), i.e.,

\[
P[\nu_n = k, \mu_n = \ell] = \frac{1}{n(n-1)}, \quad k \neq \ell
\]
\[
\quad k, \ell \in V_n.
\]

The \(\Sigma_n\)-valued rv \((\nu_n, \mu_n)\) models the randomly uniform selection of two nodes in \(V_n\) (without repetition). Each of the rvs \(\nu_n\) and \(\mu_n\) is uniformly distributed over \(V_n\). The selection rv \((\nu_n, \mu_n)\) is assumed to be independent of the random graph \(G_n\).
**Proposition 3** Under the foregoing assumptions, with \( d = 0, 1, \ldots \), the convergence

\[
\frac{N_n(d)}{n} \xrightarrow{P} L(d)
\]

holds for some scalar \( L(d) \) in \( \mathbb{R} \) if and only if

\[
\lim_{n \to \infty} \mathbb{P} [D_{n,\nu_n} = d] = L(d) \tag{2}
\]

and

\[
\lim_{n \to \infty} \text{Cov} \left[ 1 [D_{n,\nu_n} = d], 1 [D_{n,\mu_n} = d] \right] = 0. \tag{3}
\]
Remarks

- **True generalization:** Under Assumption 1, for each \( n = 1, 2, \ldots \) we have the distributional equalities
  \[
  D_{n,\nu_n} =_{st} D_{n,1} \quad \text{and} \quad D_{n,\mu_n} =_{st} D_{n,1}, \quad \text{and}
  \]
  \[
  (D_{n,\nu_n}, D_{n,\mu_n}) =_{st} (D_{n,1}, D_{n,2})
  \]

- **No** assumption on the sequence \( \{G_n, \ n = 2, 3, \ldots \} \), but also no operational ability of **equating** asymptotically the two different degree distributions available in the homogeneous case. Yet, ...

- Applies when the graphs \( \{G_n, \ n = 2, 3, \ldots \} \) are (i) deterministic (ii) **non**-homogeneous Barabási-Albert model (and other growth models) and (iii) many other models
Conclusions

- The degree distribution of a node vs. the degree distribution of the network
  - A word of caution even in homogeneous parametric networks!
- Equality in the limit under the appropriate scaling: Often true but still needs to be established
  - Theory given earlier provides a framework for doing so
- **Counterexample:** Preferential attachment vs. hidden variable models (random threshold graphs with exponential fitness)
  - **Not** comparable!
- Convergence in distribution of empirical frequencies to a non-degenerate rv
- Do not always believe what you read!

- Non-homogeneous models and random sampling
  - A new approach to convergence in growth models?