LOCAL LARGE DEVIATIONS FOR EMPIRICAL LOCALITY MEASURE OF TYPED RANDOM GRAPH MODELS

BY Kwabena Doku-Amponsah

University of Ghana

Abstract. In this article, we prove a local large deviation principle (LLDP) for the empirical locality measure of typed random networks on \( n \) nodes conditioned to have a given empirical type measure and empirical link measure. From the LLDP, we deduce a full large deviation principle for the typed random graph, and the classical Erdos-Renyi graphs, where \( nc/2 \) links are inserted at random among \( n \) nodes. No topological restrictions are required for these results.

1. Introduction and Background

1.1 Introduction

The Erdos-Renyi graph \( G(n, nc/2) \) is the simplest imaginary random, which arises by taking \( n \) nodes and inserting a fixed number \( nc/2 \) links at random among the \( n \) nodes. See, Van Hofstad [9]. Some large deviation principles for this random graph have been found. See, Bordenave & Caputo [1], Doku-Amponsah [4], and Doku-Amponsah and Moerters [7]. Doku-Amponsah and Moerters [7] provided LDPs for the near-critical or sparse typed random graphs with this model as a special case. Bordenave and Caputo [1] obtained large deviation principle for the empirical neighbourhood measure of the model \( G(n, nc/2) \). Large deviation for the empirical distribution was presented in Doku-Amponsah [4] using the method of types and the coupling argument presented by Boucheron et. al [2].

In this paper we present a Local Large deviation principle for the empirical locality measure of the conditional typed random graph model. See Bakhtin [3], Doku-Amponsah [5], Doku-Amponsah [6] for similar results for the empirical measure of iid random variables, the empirical offspring measure of multitype Galton-Watson trees and the empirical locality measure of the sparse typed random graphs, in which links appear in the graph with fix probability depending the type of the links.

1.2 The coloured random graph model.

Denote by \( Z \) a finite alphabet or type set. Let \( [n] \) be a fixed set of \( n \) nodes, say \( [n] = \{1, \ldots, n\} \). Denote by \( \mathcal{G}_n \) the set of all (simple) graphs with node set \( [n] = \{1, \ldots, n\} \) and link set \( \mathcal{E} \).

Let \( q_n : \mathcal{X} \times \mathcal{X} \to [0, 1] \) be a symmetric function and \( \sigma \) a probability measure on \( Z \). We can define the typed random graph \( Z \) with nodes \( [n] \) as follows: we Allocate to each node \( i \in [n] \) type \( Z(i) \) independently according to the type law \( \sigma \). And given the types, we connect any two nodes \( i, j \in [n] \), independently of everything else, with a link probability \( q_n(Z(i), Z(j)) \) otherwise keep them disconnected. We shall always look at \( Z = ((Z(i) : i \in [n]), \mathcal{E}) \) under the combine law of graph and type. We interpret \( Z \) as typed random graph and look at \( Z(i) \) as the type of the node \( i \). Denote by

\begin{align*}
\text{Mathematics Subject Classification :} & 94A15, 94A24, 60F10, 05C80 \\
\text{Keywords :} & \text{Local large deviation, conditional typed random graphs, relative entropy, method of types, coupling, classical Erdos-Renyi graph} \\
\text{Address :} & \text{Statistics Department, University of Ghana, Box LG 115, Legon, Ghana. E-mail: kdoku@ug.edu.gh.}
\end{align*}
\( G_n([n], Z) \) the set of all typed graphs with type set \( Z \) and \( n \) nodes. The typed random graph models may be considered in three regimes, the \textit{near-critical}, \textit{subcritical} and \textit{supercritical} cases. Thus, one may consider the cases when the link probabilities satisfy \( \alpha_n^{-1} q_n(a, b) \to \lambda(a, b) \), for all \( a, b \in X \), while the sequence \( (\alpha_n) \) is such that either \( n\alpha_n \to 1 \) or \( n\alpha_n \to 0 \) or \( n\alpha_n \to \infty \) and \( \lambda: Z \times Z \to [0, \infty) \).

By \( L(Z) \) we denote the set of all probability distributions on \( Z \), \( \tilde{L}(Z) \) we denote the set of all finite distributions on \( Z \), and by \( M(Z) \) we denote the set of counting measures on \( Z \).

For any typed graph \( Z = ((Z(i) : i \in [n]), \mathcal{E}) \) with \( n \) nodes we define a probability distribution, the \textit{empirical type distribution} \( P^1 \in L(Z) \), by

\[
P^1(a) := \frac{1}{n} \sum_{i \in [n]} \delta_{Z(i)}(a), \quad \text{for } a \in Z,
\]

and a symmetric finite distribution, the \textit{empirical link distribution} \( P^2 \in \tilde{L}(Z \times Z) \), by

\[
P^2(a, b) := \frac{1}{n} \sum_{(i, j) \in \mathcal{E}} [\delta_{Z(i), Z(j)} + \delta_{Z(j), Z(i)}](a, b), \quad \text{for } a, b \in Z.
\]

The total mass \( \|P^2\| \) of the empirical link distribution is \( 2|\mathcal{E}|/n \). Finally we define a further probability distribution, the \textit{empirical locality distribution} \( P \in L(Z \times N(Z)) \), by

\[
P(a, e) := \frac{1}{n} \sum_{i \in [n]} \delta_{Z(i), M(i)}(a, e), \quad \text{for } (a, e) \in Z \times M(Z),
\]

where \( M(i) = (m^i(b), b \in Z) \) and \( m^i(b) \) is the number of nodes of type \( b \) linked to node \( i \). For every \( p \in L(Z \times M(Z)) \) let \( p_1, p_2 \) be the \( Z \)-marginal, respectively the \( M(Z) \)-marginal, of the measure \( p \). Moreover, we define a measure \( \langle p(\cdot, e), e(\cdot) \rangle \in \tilde{L}(Z \times Z) \) by

\[
\langle p(\cdot, e), e(\cdot) \rangle(a, b) := \sum_{e \in M(Z)} p(a, e)e(b), \quad \text{for } a, b \in Z.
\]

Define the function \( \Psi: L(Z \times M(Z)) \to L(Z) \times \tilde{L}(Z \times Z) \) by \( \Psi(p) = (p_1, \langle p(\cdot, e), e(\cdot) \rangle) \). Note that \( \Psi(P) = (P^1, P^2) \), and if these quantities are defined as empirical locality, type, and link measures of a type graph.

To present the LDP, we shall call a pair of distributions \( (\pi, p) \in \tilde{L}(Z \times Z) \times L(Z \times M(Z)) \) \textit{sub-consistent} if \( \langle p(\cdot, e), e(\cdot) \rangle(a, b) \leq \pi(a, b) \), for all \( a, b \in Z \).

We shall call the triple \( (\eta, \pi, p) \in L(Z) \times \tilde{L}(Z \times Z) \times L(Z \times M(Z)) \) \textit{consistent} if

\[
\langle p(\cdot, e), e(\cdot) \rangle(a, b) \leq \pi(a, b) \text{ and } p_1 = \eta.
\]

1.3 The conditional typed random graph models. Throughout this section we may assume that \( \omega(a) > 0 \) for all \( a \in Z \). We observe that the law of the typed random graph conditioned to have a given empirical typed measure \( \eta_n \) and empirical link measure \( \pi_n \),

\[
\mathbb{Q}_{(\eta_n, \pi_n)} = \mathbb{Q}\{ \cdot \mid \Psi(M) = (\eta_n, \pi_n) \},
\]

can be described in the following manner:

- allocate types to the nodes by sampling without replacement from the collection of \( n \) types, which contains any type \( a \in Z \) exactly \( n\eta_n(a) \) times;
We write Section 3. See subsections 3.1, 3.2 and 3.3. The main Theorems of the paper; Theorem 2.1 and 2.2 and Corollary 2.3. This results are proved in The remaining part of the paper is arranged in the following way: Section 2 starts with statement of the distribution of the typed random graph

\[ n_m(a, b) := \begin{cases} n \pi_n(a, b) & \text{if } a \neq b \\ \frac{n}{2} \pi_n(a, b) & \text{if } a = b. \end{cases} \]

Theorem 2.1 (LLDP) Let \( z \)

\[ \text{Then, we have } P \Longleftrightarrow z \text{ for any functional } q. \]

\[ \text{Thus, } \exists \text{ a weak neighborhood } B_{\eta} \text{ such that } \]

\[ Q_{\eta, \pi_n}(z) = Q_{\eta, \pi_n}(Z = z) = Q(Z = z) \Psi(P) = (\eta, \pi_n) \]

the distribution of the typed random graph \( y \) conditioned to have typed law and link law \( (\eta, \pi_n) \) respectively.

\[ J_{\eta, \pi}(p) = \begin{cases} H(p \parallel q) & \text{if } (\pi, p) \text{ is sub-consistent and } p_1 = \eta \\ \infty & \text{otherwise.} \end{cases} \quad (2.1) \]

where

\[ q(a, e) = \eta(a) \prod_{b \in \mathbb{Z}} \frac{e^{-\pi(a,b)/\eta(a)} \pi(a,b)/\eta(a)^e(b)}{e(b)!}, \text{ for } e \in \mathcal{M}(\mathbb{Z}). \]

**Theorem 2.1** (LLDP). Suppose the sequence \((\eta_n, \pi_n)\) converges to \((\pi, \eta) \in \mathcal{L}(\mathbb{Z}) \times \mathcal{L}_s(\mathbb{Z} \times \mathbb{Z})\). Let \( z = \{(z(i), i \in [n]), \mathcal{E}\} \) be coloured random graph conditioned on the event \( \{\Psi(P) = (\eta_n, \pi_n)\} \). Then, we have

(i) for any functional \( p \in \mathcal{L}(\mathbb{Z} \times \mathcal{M}(\mathbb{Z})) \) and a number \( \varepsilon > 0 \), there exists a weak neighborhood \( B_p \) such that

\[ Q_{\eta, \pi_n}(z \in \mathcal{G}([n], (\eta_n, \pi_n)) \mid P_z \in B_p) \leq e^{-n J_{\eta, \pi}(\eta, \pi_n) - n\varepsilon + o(n)}. \]

(ii) for any \( p \in \mathcal{L}_s(\mathbb{Z} \times \mathbb{Z}_s) \), a number \( \varepsilon > 0 \) and a fine neighborhood \( B_p \) we have the asymptotic estimate:

\[ Q_{\eta, \pi_n}(z \in \mathcal{G}([n], (\eta_n, \pi_n)) \mid P_z \in B_p) \geq e^{-n J_{\eta, \pi}(\eta, \pi_n) + n\varepsilon - o(n)}. \]

Finally, we state in Theorem 2.2 the full LDP for the Typed random graph.

**Theorem 2.2** (LDP). Suppose the sequence \((\eta_n, \pi_n)\) converges to \((\pi, \eta) \in \mathcal{L}(\mathbb{Z}) \times \mathcal{L}_s(\mathbb{Z} \times \mathbb{Z})\). Let \( z = \{(z(i), i \in [n]), \mathcal{E}\} \) be coloured random graph conditioned on the event \( \{\Psi(P) = (\eta_n, \pi_n)\} \).

(i) Let \( F \) be open subset of \( \mathcal{L}(\mathbb{Z} \times \mathbb{Z}_s) \). Then we have

\[ \limsup_{n \to \infty} \frac{1}{n} \log Q_{\eta, \pi_n}(z \in \mathcal{G}([n], (\eta_n, \pi_n)) \mid P_z \in F) \leq - \inf_{p \in F} J_{\eta, \pi}(p). \]

(ii) Let \( \Gamma \) be closed subset of \( \mathcal{L}(\mathbb{Z} \times \mathbb{Z}_s) \). Then we have

\[ \liminf_{n \to \infty} \frac{1}{n} \log Q_{\eta, \pi_n}(y \in \mathcal{G}([n], (\eta_n, \pi_n)) \mid P_z \in \Gamma) \geq - \inf_{p \in \Gamma} J_{\eta, \pi}(p). \]
Corollary 2.3 (LDP). Suppose $D_z$ is the degree distribution of $z \in G([n], nc/2)$ and let $\Gamma$ be subset of $\mathcal{L}(\mathbb{Z} \times \mathbb{Z}_*)$. Then we have

\[
- \inf_{p \in \text{int}(\Gamma)} I_c(p) \leq \liminf_{n \to \infty} \frac{1}{n} \log Q_{(\eta_n, \pi_n)} \left\{ z \in G([n], nc/2) \mid D_z \in \Gamma \right\} \\
\leq \limsup_{n \to \infty} \frac{1}{n} \log Q_{(\eta_n, \pi_n)} \left\{ z \in G([n], nc/2) \mid D_z \in \Gamma \right\} \leq - \inf_{p \in \partial(\Gamma)} I_c(p),
\]

where $q_c$ is the poisson distribution with mean $c$, $I_c(p) = H(p \mid q_c)$ and $\infty$ if otherwise.

3. Proof Main Results

Lemma 3.1 (Doku-Amponsah [4]). For any empirical locality measure $p_n$, we have

\[
e^{-nH(p_n \mid q_n) - o(n)} \leq Q_{(\eta_n, \pi_n)} \{ P = p_n \} \leq e^{-nH(p_n \mid q_n) + o(n)},
\]

where

\[
q_n(a, e) = \eta_n(a) \prod_{b \in \mathbb{Z}} \frac{e^{-\pi(a,b)/\eta_n(a)} [\pi(a,b)/\eta_n(a)]^e(b)}{e(b)!}, \text{ for } e \in \mathcal{M}(\mathbb{Z}).
\]

Note that $q_n \to q$ where

\[
q(a, e) = \eta(a) \prod_{b \in \mathbb{Z}} \frac{e^{-\pi(a,b)/\eta(a)} [\pi(a,b)/\eta(a)]^e(b)}{e(b)!}, \text{ for } e \in \mathcal{M}(\mathbb{Z}).
\]

3.1 Proof of Theorem 2.1

We denote by $\tilde{z}$ the random allocation process and note that for any consistent triple $(\eta_n, \pi_n, p_n)$ we have

\[
\frac{Q_{(\eta_n, \pi_n)} \{ Z = z \mid P_z \in B_p \}}{\tilde{Q}_{(\eta_n, \pi_n)} \{ Z = z \mid P_z \in B_{\tilde{p}} \}} = \frac{\tilde{z} \{ y : P_z = p_n \}}{\tilde{z} \{ z : P_z = \tilde{p}_n \}} = \frac{\tilde{z} \{ z : (P^1(z), P^2(z)) = (\eta_n, \pi_n) \}}{\tilde{z} \{ z : P_z = \tilde{p}_n \}}.
\]

(3.1)

Now we define a locality of the functional $\rho$ as follows:

\[
B_p = \left\{ \mu \in \mathcal{L}(\mathbb{Z} \times \mathbb{Z}_*) : \langle \mu, \log \frac{\mu}{\eta} \rangle > \langle p, \log \frac{p}{q} \rangle - \frac{\varepsilon}{2} \right\},
\]

Using [7, Lemma 3.1], under the condition $P_z \in B_p$ and $\tilde{p} = q$ we have that

\[
e^{-nH(p_n \mid q_n) + nH(q \mid q_n) - o(n)} \leq \frac{dQ_{(\eta_n, \pi_n)}(z)}{d\tilde{Q}_{(\eta_n, \pi_n)}(z)} \leq e^{-nH(p_n \mid q_n) + nH(q \mid q_n) + o(n)},
\]

(3.2)

Fix $\varepsilon > 0$, and note that by 3.2 we have

\[
Q_{(\eta_n, \pi_n)} \{ z \in G([n], (\eta_n, \pi_n)) \mid P_z \in B_p \} \leq \int_{P_z \in B_p} e^{-nH(p_n \mid q_n) + nH(q \mid q_n) + o(n)} d\tilde{Q}_{(\eta_n, \pi_n)}(z) \leq e^{-nJ(\eta, \pi)(p) - n\varepsilon + o(n)}.
\]

Also we have
\[
Q(\eta_n, \pi_n) \left\{ z \in G([n], (\eta_n, \pi_n)) \left| P_z \in B_p \right. \right\} \geq \int_{P_z \in B_p} e^{-nH(p_n \| q_n) + nH(q \| q_n) - o(n)} d\tilde{Q}(\eta_n, \pi_n)(z)
\]

which completes the proof of the Theorem.

The proof of Theorem 2.2 below, follows from Theorem 2.1 above using similar arguments as in [3, p. 544].

### 3.2 Proof of Theorem 2.2

**Proof.** We observe that the empirical locality distribution is a probability distribution and so it contain in the unit ball in \( \hat{L}(\mathcal{Z} \times \mathcal{Z}^*) \). Henceforth, without loss of generality we can assume that the set \( \Gamma \) in Theorem 2.2(ii) above is relatively compact. If we choose any \( \varepsilon > 0 \), then for each functional \( p \in \Gamma \) we can find a weak neighbourhood such that the estimate of Theorem 2.1(i) above holds. From all these neighbourhood, we choose a finite cover of \( G([n], (\eta_n, \pi_n)) \) and sum up over the estimate in Theorem 2.1(i) above to obtain

\[
\lim_{n \to \infty} \frac{1}{n} \log Q(\eta_n, \pi_n) \left\{ z \in G([n], (\eta_n, \pi_n)) \left| P_z \in \Gamma \right. \right\} \leq - \inf_{p \in \Gamma} J(\eta, \pi)(p) + \varepsilon.
\]

Since \( \varepsilon \) was arbitrarily chosen and the lower bound in Theorem 2.1(ii) is implies the lower bound in Theorem 2.2(i) we have the required results which completes the proof.

### 3.3 Proof of Corollary 2.3

**Proof.** The proof of Corollary 2.3 follows from Theorem 2.2 if we take \( P = D_z \) then \( \Psi(D_z) = 2|\mathcal{E}|/n = c \). Hence, in the case of the classical Erdos-Renyi graph \( G([n], nc/2) \), the rate function \( J(\eta, \pi)(p) \) reduces to \( H(p \| q_n) \) and \( \infty \) if otherwise.

**Acknowledgement**

This article was finalized at the Carnigie Banga-Africa, June 27-July 2017 writeshop, in Koforidua.

**References**

[1] C. Bordenave and P. Caputo Large deviations of empirical neighbourhood distribution in sparse random graphs. *arxiv:1308.5725(2013).*

[2] S. Boucheron, F. Gamboa and C. Leonard. Bins and balls: Large deviations of the empirical occupancy process. *Ann. Appl. Probab. 12 607-636 (2002).*

[3] I.V. Bakhtin. Spectral Potential, Kullback Action, and Large deviations of empirical measureson measurable spaces. *Theory of Probability and application. Vol. 50,No.4.(2015) pp.535-544.*

[4] K. Doku-Amponsah. Exponential Approximation of empirical neighbourhood measures of random graphs by random allocation. *International Journal of Statistics and Probability, Vol. 3, No 2; 2014. pp 110-120.*

[5] K. Doku-Amponsah. Local Large deviations: MCMillian Theorem for multitype Galton-Watson Processes.
[6] K. Doku-Amponsah. Local Large deviations: a McMillian Theorem for Typed Random Graph Processes.

[7] K. Doku-Amponsah and P. Mörters. Large deviation principle for empirical measures of coloured random graphs. *The Annals of Applied Probability, 20*(6), 1989-2021.

[8] A. Dembo and O. Zeitouni. Large deviations techniques and applications. Springer, New York, (1998).

[9] R. Van Der Hofstad Random Graphs and Complex Networks. *Eindhoven University of Technology*. Unpublished Manuscript.