# Temporal Statistical Analysis of Degree Distributions in an Undirected Landline Phone Call Network Graph Series

**Orgeta Gjermëni** [ID]

Department of Mathematics, University Ismail Qemali, Str. Kosova, 9400 Vlore, Albania; o.gjermeni@gmail.com

**Abstract:** This article aims to provide new results about the intraday degree sequence distribution considering phone call network graph evolution in time. More specifically, it tackles the following problem. Given a large amount of landline phone call data records, what is the best way to summarize the distinct number of calling partners per client per day? In order to answer this question, a series of undirected phone call network graphs is constructed based on data from a local telecommunication source in Albania. All network graphs of the series are simplified. Further, a longitudinal temporal study is made on this network graphs series related to the degree distributions. Power law and log-normal distribution fittings on the degree sequence are compared on each of the network graphs of the series. The maximum likelihood method is used to estimate the parameters of the distributions, and a Kolmogorov–Smirnov test associated with a *p*-value is used to define the plausible models. A direct distribution comparison is made through a Vuong test in the case that both distributions are plausible. Another goal was to describe the parameters' distributions' shape. A Shapiro-Wilk test is used to test the normality of the data, and measures of shape are used to define the distributions' shape. Study findings suggested that log-normal distribution models better the intraday degree sequence data of the network graphs. It is not possible to say that the distributions of log-normal parameters are normal.

**Keywords:** longitudinal; degree distribution; network graph; phone call data; power law; log-normal

## 1. Introduction

Most studies related to phone call network graphs are based on mobile call data [1–7] rather than on landline phone call data [8–11]. Network graphs are seen as static, and rarely [11,12] are they pursued in temporal studies. A local telecommunicating data set from Albania was used in [10] to construct a static network graph. The tails of the empirical distributions were analyzed on the greatest connected component related to: the number of phone calls per client, the total duration of calls per client in seconds, and the distinct number of calling partners per client. The network graph was considered in both cases, directed and not directed. A comparison between power law (PL) and log-normal (LN) fit was made in the tail of the distributions, but it could not be concluded which of them had a determinate dominance over the other. Tail analysis in vertex degree or vertex strength distribution in communication network graphs is important because it gives information about hubs and rare events. Hubs are highly connected vertices, which are hypothesized to act as focal points for the convergence or divergence of information.

Considering the network graph as static, and the concentration only at its greatest connected component may have influenced our findings in [10]. This study aims to provide new results about the intraday degree sequence distribution considering phone call network graph evolution in time.

Phone call communication relations have a survival time. Network graph evolution in time is related to the network graph's topology state, which is in a continuous change. A day's snapshot is used to show the topology state of the network graph in a time point.

This article tackles the following problem. Given a large amount of landline phone call data records, what is the best way to summarize the distinct number of calling partners per client per day? In order to answer this question, I construct an undirected phone call network graphs series, with all network graphs of the series simplified. Further, a longitudinal temporal study is made on this network graph series related to the degree distributions. PL and LN are compared on each of the degree sequences of the network graph series.

The vertex degree is related to the number of distinct callers and the number of distinct subjects that are called by an active phone client. This relation is conditioned by the fact that the network graphs are undirected and simplified. The analysis aimed to determine the distribution that yielded a better fit to model the data related to the degree sequence in each time step. It is shown that the LN model is better, mainly because it covers a large amount of data, and it was determined by the tests to be more reliable than the PL model. I also considered the distributions' shape of the LN parameters and described them. The results show that the distributions of LN parameters were not normal.

## 2. Materials and Methods

### 2.1. Data Preparation

The data set is provided by a local telecommunicating operator positioned in the south of Albania, which covers approximately 4% of the landline market in the country. Clients' identities were substituted with numbers to conserve privacy (see Supplementary Material). The study is based only on phone calls inside the operator's client network, and not outside it. The reason for this restriction is based on the evidence that phone number data which did not belong to the operator would be incomplete.

Phone calls took place in November 2014. On 28 November, Albania celebrates Independence Day, and on the 29th, Liberation Day. From a total of 81,591 phone calls, 41, which were without call durations, and 7442, which lasted less than 10 s, were excluded from the study. The reason for this exclusion is that these calls were lost calls or wrong numbers and might have affected the accuracy of the results. Thus, the total data set used for the study was 90.83% of the initial data set. Active clients are considered only those that were engaged in at least in one phone call (made or received) that lasted at least 10 s, amounting to a total number of 3287. Multiple phone call relations between any two clients were treated as single phone call relations. This statistical technique, about filtering and extracting the best sample that would reflect the global calling patterns related to the number of calling partners per client, has been applied by other authors in telecommunication data [1,2].

Degree distribution in the communication system was studied by observing 30 network graphs, which were constructed by splitting the data set for each day of the month. The network graphs are denoted by $G_i = (V_i, E_i)$. The vertex set (active phone clients) is $V_i$, and the edge set is $E_i$ ($G_1$ is the network graph of the first day of the month, $G_2$ for the second day, and so on). Each edge represents a communication relation between two phone clients. Thus, if $v_1$ and $v_2$ are vertices, then an undirected edge $(v_1, v_2)$ is between them only if $v_1$ has made or received at least one phone call from $v_2$ or the reverse. Multiple relations between two vertices are simplified as only one edge. In Table 1, the topology techniques various authors have used are mentioned. The table includes the following information: the type of telecommunication data, the time interval, the relation's direction, the relation's mutuality, the relation's simplification, and the relation's weight. There is no precise topology technique on how to treat mobile or landline data. Variability depends on the goal of the scientific research.

$G_1, G_2, \ldots, G_{30}$ is defined as the temporal network graph series. The network graph $G_i$ is constructed based only on the data of the $i$-th day. Vertex degree [13] in a network graph is defined

as the number of edges incident on that vertex. Let $d_v$ denote the degree of the vertex $v$ and, with $\{d_v^i\}_{v \in V_i}$, the vertex degree sequence of $G_i$. The fraction of vertices $v$ that have $d_v = x$ is denoted by $p_x$. This can also be interpreted as—the probability that a vertex chosen uniformly at random has a degree equal to $x$. The set of $\{p_x\}_{x \geq 0}$ defines the degree distribution of the network graph.

**Table 1.** A general overview of topology statistical techniques used to analyze phone call data. Abbreviations: m: months; w: weeks; d: days; -: not applicable.

| Authors | Data | Time | Directed | Mutual | Simplified | Weighted |
|---|---|---|---|---|---|---|
| Nanavati et al. [1,2] | mobile | 1 w, 1 m | yes | no | yes | no |
| Seshadri et al. [3] | mobile | 2 m | no | yes | yes | yes |
| Dong et al. [4] | mobile | 1 m | no | no | yes | no |
| Onnela et al. [6,7] | mobile | 18 w | no | Both (yes, no) | - | yes |
| Ye et al. [12] | mobile | 10 d | yes | no | no | no |
| Noka (Jani) & Hoxha [5] | mobile (calls, SMS) | 1 m | no | yes | - | yes |
| Aiello et al. [8,9] | landline | 1 d | yes | no | no | no |
| Gjermëni & Ramosaco [10] | landline | 1 m | Both (yes, no) | no | yes | yes |

## 2.2. Temporal Statistical Analysis

At first, for each of the network graphs of the series $\{G_i\}_1^{30}$, the vertex degree sequence $\{d_v^i\}_{v \in V_i}$ was computed. The normality of $\{d_v^i\}_{v \in V_i}$ was controlled. Thus, a histogram and Q–Q plot were constructed. The Shapiro-Wilk test was performed on the degree sequence, and the basic statistics were calculated. If the $p$-value of the test [14–16] was less than chosen alpha level 0.05, it was considered as evidence that the data did not come from a normally distributed population.

Skewness [17] and kurtosis were used to determine whether the empirical distribution was heavy-tailed. Increasing kurtosis was associated with the "movement of probability mass from the shoulders of a distribution into its centre and tails" [18]. Leptokurtic distributions (kurtosis values are greater than 3) partly comprise heavy-tailed distributions [19]. Probability distribution functions that decay slower than an exponential are called heavy-tailed distributions. According to [20], a distribution is heavy-tailed if and only if its tail function is a heavy-tailed function. A non-negative function is said to be heavy-tailed if it fails to be bound by a decreasing exponential function.

PL and LN distributions are heavy-tailed. These distributions are chosen to be fitted on data for $x \geq x_{min}$, because it is not always possible to get a good fitting for all the data. A random variable $X$ follows a PL distribution for $X \geq x_{min}$ if its probability mass function $P(X \geq x_{min})$ is

$$f(x, x_{min}) = \frac{x^{-\alpha}}{\zeta(\alpha, x_{min})}$$

where $\zeta(\alpha, x_{min}) = \sum_{x=x_{min}}^{\infty} x^{-\alpha}$ is the general $\zeta$−Riemann function. $\alpha$ is the scaling parameter of the distribution. A random variable $X$ follows a LN distribution for $X \geq x_{min}$ if

$$f(x, x_{min}) = \frac{2}{\sqrt{2\pi}\sigma x} \left[ \text{erfc}\left( \frac{\ln x_{min} - \mu}{\sqrt{2}\sigma} \right) \right]^{-1} \exp\left[ -\frac{(\ln x - \mu)^2}{2\sigma^2} \right].$$

$\mu$ and $\sigma$ are parameters of the distribution. The estimation procedure is based on the maximum likelihood method [21,22]. This technique is also applied by other authors [23].

The Kolmogorov–Smirnov statistic (KS) is used to determine goodness-of-fit, and the $p$-value based on 2500 instances of bootstrapping is computed for each of the fittings. Small KS values, and $p > 0.1$ suggest that the fitted distribution is a plausible one for the set of the data, such that $x \geq x_{min}$. If $p \leq 0.1$, then it is said that the data does not come from either a PL or an LN distribution. A reliable $p$-value is obtained when the number of data in the tail of the distribution, $n_{tail}$, is greater than 100 for PL and greater than 300 for LN [21,22].

When both PL and LN are plausible models for the data, a Vuong log likelihood test [24] between them is computed. The sign of the log likelihood ratio, $\mathcal{R}$, can be reliably used to determine which of

the models is better than the other if the *p*-value is less than 0.1. Otherwise, both models are considered equally plausible.

After that, a box plot description of temporal change on the estimated parameters of the distributions $\alpha_i$, $\mu_i$, $\sigma_i$ and their estimated $x^i_{min}$ for $i = \overline{1,30}$ is constructed. Three cases are considered:

- Case 1: the fitting made from 1;
- Case 2: the fitting made from the estimated $x_{min}$ of each distribution;
- Case 3: the fitting made from the $\min\left(x^{PL}_{min}, x^{LN}_{min}\right)$, where both distributions are plausible.

Furthermore, a shape description of parameter distributions of the best-fitted degree distribution models is made. A visualization of the log–log plots of the complementary cumulative distribution function (CCDF) ($P(d_v \geq x)$) is provided for Case 1, 2, and 3 at $G_1$ and $G_{30}$.

The statistical computation related to these distributions are made based on the following packages in the R statistical computation platform [25]: poweRlaw [26], fBasics [27], igraphdata [28], and igraph [29].

## 3. Results

After constructing the undirected landline phone call network graphs series, each of the network graphs of the series is simplified. Further, the data set I analyze here is the degree sequence $\left\{d^i_v\right\}_{v \in V_i}$ of each $G_i$ where $i = \overline{1,30}$. In this section, the results of the study are presented. They are divided into two subsections.

### 3.1. Descriptive Analysis of Degree Values

In Figure 1 and Table 2, an illustration for the case of $G_1$ and $G_{30}$ related to the histogram and Q–Q plot and the basic statistics for the set of degree values are shown. For all $G_i$, based on the Q–Q plots, a strong deviation from the straight line can be seen. Moreover, after running the Shapiro-Wilk normality test in each of them, the *p*-values were always less than $2.2 \times 10^{-16} < 0.05$. This means that the data did not come from a normal distribution.
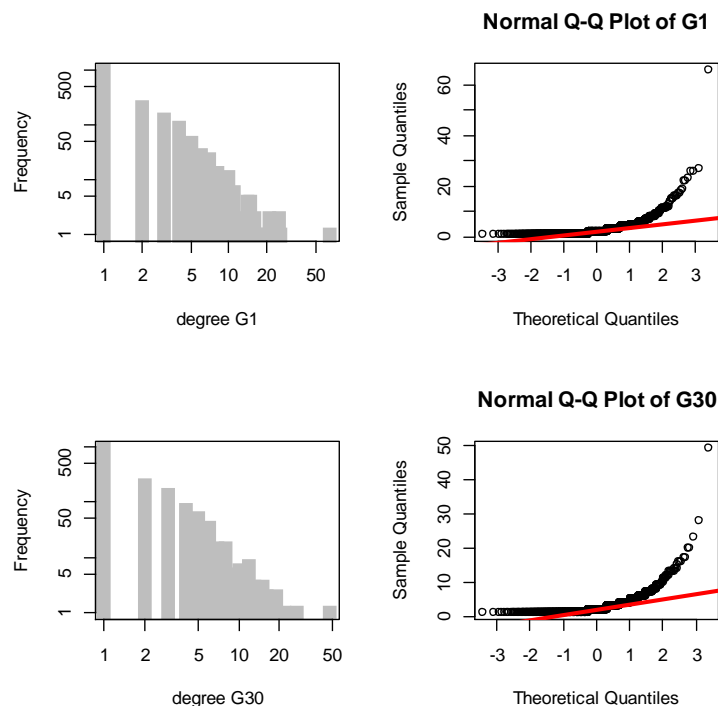


**Figure 1.** These are the histograms and the Q–Q plots of degree values of $G_1$ and $G_{30}$. Histogram axes are logarithmic.

**Table 2.** A summary of some basic statistics for the set of degree values of $G_1$ and $G_{30}$.

| Basic Stats | $G_1$ | $G_{30}$ |
|---|---|---|
| Minimum | 1 | 1 |
| First Quartile | 1 | 1 |
| Median | 2 | 2 |
| Third Quartile | 3 | 3 |
| Maximum | 66 | 49 |
| Mean | 2.79 | 2.77 |
| Skewness | 7.25 | 5.77 |
| Kurtosis | 108.41 | 57.91 |
| Mode | 1 | 1 |

Furthermore, for all $G_i$,

- the degree sequence is unimodal;
- the mode is 1;
- mean > median > mode;
- the peak of the data is on the left and the right tail is longer;
- skewness is greater than 1;
- kurtosis is greater than 3.

This means that, in all degree sequences $\left\{ d_v^i \right\}_{v \in V_i}$, a highly right (positively) skewed distribution is present. They are leptokurtic and heavy-tailed.

*3.2. Statistical Analysis of Fitted Distributions*

The results of the estimated $\alpha$ and $x_{min}$ for the PL distribution are given in Table 3. Information about the total number of data $n$ in $\left\{ d_v^i \right\}_{v \in V_i}$, as well as the quantity of the data that are in the tail of the distribution $n_{tail}$, is also given. Based on $p$-values where $p \leq 0.10$, PL is rejected only once out of 30 ($G_{20}$). The $p$-value is not reliable in six cases ($G_7$, $G_{10}$, $G_{12}$, $G_{15}$, $G_{16}$, $G_{26}$), since $n_{tail} < 100$.

The results for the estimated parameters, $\sigma$, and $x_{min}$ for the LN, information about the total number of data $n$ in $\left\{ d_v^i \right\}_{v \in V_i}$, and the quantity of data in the tail of the distribution $n_{tail}$ is given in Table 4. Based on $p$-values ($p \leq 0.10$), LN is rejected three times out of 30 ($G_2$, $G_{12}$, $G_{21}$). The $p$-value is always reliable, since $n_{tail} > 300$ in all the cases. For each $G_i$, the KS value of LN is always lower than the corresponding KS value of PL.

**Table 3.** Parameter estimates and the goodness-of-fit (KS) for the power law (PL) distribution.

| $G_i$ | $n$ | PL | | |
|---|---|---|---|---|
| | | $x_{min}$ $(n_{tail})$ | $\alpha$ | KS $(p)$ |
| 1 | 1597 | 5 (245) | 3.27 | 0.02 (0.665) |
| 2 | 1428 | 6 (153) | 3.95 | 0.04 (0.274) |
| 3 | 1561 | 5 (276) | 3.07 | 0.03 (0.195) |
| 4 | 1534 | 8 (108) | 3.78 | 0.03 (0.813) |
| 5 | 1522 | 7 (117) | 3.64 | 0.03 (0.459) |
| 6 | 1530 | 7 (132) | 3.49 | 0.03 (0.509) |
| 7 | 1531 | 9 (74) | 3.66 | 0.02 (0.895) |
| 8 | 1560 | 6 (199) | 3.37 | 0.03 (0.388) |
| 9 | 1487 | 7 (110) | 3.64 | 0.03 (0.801) |
| 10 | 1564 | 9 (76) | 3.69 | 0.03 (0.846) |
| 11 | 1545 | 7 (131) | 3.67 | 0.03 (0.819) |
| 12 | 1531 | 8 (97) | 3.46 | 0.04 (0.274) |
| 13 | 1547 | 7 (121) | 3.58 | 0.04 (0.359) |
| 14 | 1552 | 7 (133) | 3.69 | 0.02 (0.988) |
| 15 | 1555 | 9 (79) | 3.88 | 0.04 (0.569) |

**Table 3.** *Cont.*

| $G_i$ | $n$ | PL | | |
|---|---|---|---|---|
| | | $x_{min}$ ($n_{tail}$) | $\alpha$ | KS ($p$) |
| 16 | 1554 | 8 (90) | 4.11 | 0.02 (0.992) |
| 17 | 1555 | 6 (195) | 3.43 | 0.03 (0.483) |
| 18 | 1560 | 7 (126) | 3.42 | 0.03 (0.653) |
| 19 | 1494 | 5 (235) | 3.24 | 0.03 (0.204) |
| 20 | 1568 | 5 (257) | 3.31 | 0.04 (0.062) |
| 21 | 1523 | 7 (137) | 3.61 | 0.04 (0.192) |
| 22 | 1505 | 7 (130) | 3.49 | 0.03 (0.489) |
| 23 | 1479 | 7 (117) | 3.87 | 0.04 (0.377) |
| 24 | 1512 | 5 (247) | 3.14 | 0.02 (0.682) |
| 25 | 1559 | 5 (260) | 3.15 | 0.03 (0.204) |
| 26 | 1545 | 10 (60) | 3.78 | 0.03 (0.702) |
| 27 | 1530 | 6 (163) | 3.35 | 0.04 (0.130) |
| 28 | 1505 | 6 (176) | 3.18 | 0.04 (0.253) |
| 29 | 1524 | 7 (127) | 3.49 | 0.03 (0.528) |
| 30 | 1472 | 6 (151) | 3.58 | 0.02 (0.819) |

**Table 4.** Parameter estimates and the goodness-of-fit (KS) for the log-normal distribution.

| $G_i$ | $n$ | LN | | | |
|---|---|---|---|---|---|
| | | $x_{min}$ ($n_{tail}$) | $\mu$ | $\sigma$ | KS ($p$) |
| 1 | 1597 | 1 (1597) | 0.48 | 0.93 | 0.01 (0.196) |
| 2 | 1428 | 1 (1428) | 0.52 | 0.88 | 0.01 (0.016) |
| 3 | 1561 | 3 (619) | 0.38 | 0.99 | 0.01 (0.793) |
| 4 | 1534 | 2 (906) | 0.83 | 0.84 | 0.01 (0.239) |
| 5 | 1522 | 2 (913) | 0.79 | 0.81 | 0.01 (0.426) |
| 6 | 1530 | 1 (1530) | 0.54 | 0.93 | 0.01 (0.184) |
| 7 | 1531 | 4 (400) | 0.12 | 1.03 | 0.01 (0.703) |
| 8 | 1560 | 1 (1560) | 0.56 | 0.95 | 0.01 (0.251) |
| 9 | 1487 | 1 (1487) | 0.52 | 0.90 | 0.00 (0.948) |
| 10 | 1564 | 1 (1564) | 0.50 | 0.95 | 0.00 (0.876) |
| 11 | 1545 | 2 (932) | 0.71 | 0.86 | 0.01 (0.424) |
| 12 | 1531 | 1 (1531) | 0.49 | 0.96 | 0.01 (0.037) |
| 13 | 1547 | 1 (1547) | 0.53 | 0.92 | 0.01 (0.653) |
| 14 | 1552 | 2 (949) | 0.84 | 0.81 | 0.01 (0.247) |
| 15 | 1555 | 1 (1555) | 0.47 | 0.97 | 0.00 (0.969) |
| 16 | 1554 | 2 (910) | 0.78 | 0.81 | 0.01 (0.556) |
| 17 | 1555 | 2 (926) | 0.72 | 0.88 | 0.01 (0.413) |
| 18 | 1560 | 4 (397) | −0.52 | 1.16 | 0.01 (0.895) |
| 19 | 1494 | 1 (1494) | 0.51 | 0.92 | 0.00 (0.898) |
| 20 | 1568 | 1 (1568) | 0.56 | 0.90 | 0.01 (0.162) |
| 21 | 1523 | 1 (1523) | 0.57 | 0.93 | 0.02 (0.009) |
| 22 | 1505 | 2 (894) | 0.74 | 0.87 | 0.01 (0.464) |
| 23 | 1479 | 2 (859) | 0.81 | 0.80 | 0.01 (0.223) |
| 24 | 1512 | 1 (1512) | 0.5 | 0.95 | 0.01 (0.622) |
| 25 | 1559 | 1 (1559) | 0.48 | 0.96 | 0.01 (0.379) |
| 26 | 1545 | 1 (1545) | 0.47 | 0.99 | 0.01 (0.467) |
| 27 | 1530 | 3 (577) | 0.10 | 1.03 | 0.01 (0.871) |
| 28 | 1505 | 1 (1505) | 0.52 | 0.95 | 0.01 (0.590) |
| 29 | 1524 | 2 (917) | 0.66 | 0.89 | 0.01 (0.740) |
| 30 | 1472 | 2 (882) | 0.74 | 0.80 | 0.01 (0.488) |

Information about the LN and PL distribution, for cases where both are plausible, is shown in Table 5. Since $min\left(x_{min}^{PL}, x_{min}^{LN}\right) = x_{min}^{PL}$, it can be said that LN is conditioned by PL. The sign of the log likelihood ratio, $\mathcal{R}$ does not reliably determine which of the models is better than the other, because $p$-value is not less than 0.1. In this way, both models are considered equally plausible. "-" denotes cases where no comparison between LN and PL can be made, because they are defined as not plausible when $x \geq min\left(x_{min}^{PL}, x_{min}^{LN}\right)$ ($G_2$, $G_{12}$, $G_{20}$, $G_{21}$).

**Table 5.** Results of the Vuong log likelihood ratio test $\mathcal{R}$ for LN and PL.

| $G_i$ | LN Conditioned by PL | | | LN vs. PL | |
|---|---|---|---|---|---|
| | $x_{min}$ ($n_{tail}$) | $\mu$ | $\sigma$ | $\mathcal{R}$ | $p$ |
| 1 | 5 (245) | −3.36 | 1.59 | 0.516 | 0.606 |
| 2 | - | - | - | - | - |
| 3 | 5 (276) | −0.02 | 1.08 | 1.232 | 0.218 |
| 4 | 8 (108) | −490.96 | 13.36 | 0.268 | 0.788 |
| 5 | 7 (117) | −48.44 | 4.40 | 0.087 | 0.931 |
| 6 | 7 (132) | −5.14 | 1.77 | 0.328 | 0.743 |
| 7 | 9 (74) | −10.23 | 2.22 | 0.114 | 0.909 |
| 8 | 6 (199) | −7.20 | 2.03 | 0.253 | 0.8 |
| 9 | 7 (110) | −505.22 | 13.89 | -0.007 | 0.994 |
| 10 | 9 (76) | −512.04 | 13.85 | 0.424 | 0.671 |
| 11 | 7 (131) | −519.53 | 14.02 | 0.459 | 0.647 |
| 12 | - | - | - | - | - |
| 13 | 7 (121) | −123.13 | 7.00 | 0.342 | 0.732 |
| 14 | 7 (133) | −266.38 | 10.02 | 0.052 | 0.958 |
| 15 | 9 (79) | −20.99 | 2.88 | 0.063 | 0.95 |
| 16 | 8 (90) | −440.36 | 11.97 | −0.032 | 0.975 |
| 17 | 6 (195) | −2.79 | 1.48 | 0.447 | 0.655 |
| 18 | 7 (126) | −1.35 | 1.29 | 0.458 | 0.647 |
| 19 | 5 (235) | −0.21 | 1.07 | 0.928 | 0.353 |
| 20 | - | - | - | - | - |
| 21 | - | - | - | - | - |
| 22 | 7 (130) | −2.56 | 1.45 | 0.361 | 0.718 |
| 23 | 7 (117) | −5.83 | 1.71 | 0.243 | 0.808 |
| 24 | 5 (247) | −6.89 | 2.09 | 0.399 | 0.69 |
| 25 | 5 (260) | −0.66 | 1.19 | 0.855 | 0.393 |
| 26 | 10 (60) | −490.04 | 13.32 | 0.081 | 0.936 |
| 27 | 6 (163) | −3.12 | 1.55 | 0.0378 | 0.705 |
| 28 | 6 (176) | −0.26 | 1.14 | 0.772 | 0.44 |
| 29 | 7 (127) | −6.08 | 1.88 | 0.185 | 0.85 |
| 30 | 6 (151) | −2.60 | 1.40 | 0.411 | 0.68 |

Box plots of the temporal change on the estimated parameters of the distributions $\alpha_i$, $\mu_i$, $\sigma_i$ and their estimated $x_{min}^i$ for $i = \overline{1,30}$ are shown in Figure 2. Case 1 is based on box plots $LN_1$ and $PL_1$; Case 2 is based on box plots $LN_{x_{min}}$ and $PL_{x_{min}}$; Case 3 is based on box plots $LN_{x_{min-PL}}$.
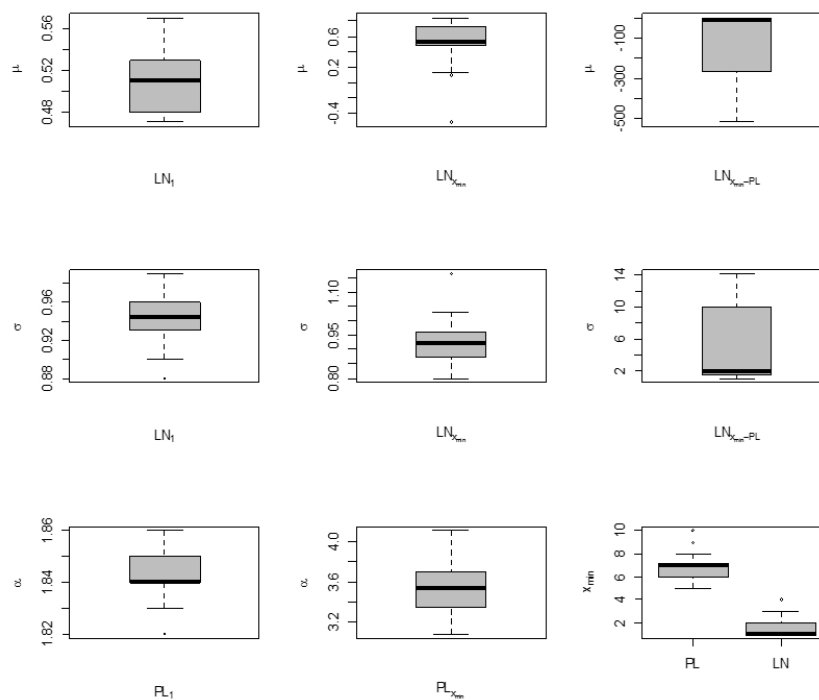


**Figure 2.** Box plots of temporal changes of $x_{min}$ and parameters $\alpha$, $\mu$, $\sigma$ of the distributions.

It was found that between weekdays, weekends, and holidays, there was no substantial change related to the degree sequence of the network graph. The PL was not rejected in either of the weekends or holidays, but it was not reliable for one of the weekends ($G_{16}$, $G_{17}$). LN was rejected for one of the weekend days ($G_2$) but was otherwise always reliable. For weekdays, PL and LN models were both rejected and accepted.

Some statistics about the temporal change of LN parameters are given in Table 6. A shape description of the LN parameter distribution is given as follows:

1. $\mu$: In all three cases, $\mu$ does not come from a normal distribution, and its shape is described as follows:

   - Case 1: approximately symmetric and platykurtic;
   - Case 2: highly negative skewed and leptokurtic;
   - Case 3: highly negative skewed and platykurtic.

2. $\sigma$: In Cases 1 and 2, based on the Shapiro-Wilk test, the normal distribution was not rejected, but it was in Case 3. The shape of the $\sigma$ distribution is described as follows:

   - Case 1: approximately symmetric and platykurtic;
   - Case 2: moderately skewed and platykurtic;
   - Case 3: moderately skewed and platykurtic.

**Table 6.** A summary of statistics related on the temporal change of LN parameters.

| Statistics | $\mu$ | | | $\sigma$ | | |
|---|---|---|---|---|---|---|
| | $LN_1$ | $LN_{x_{min}}$ | $LN_{x_{min}-PL}$ | $LN_1$ | $LN_{x_{min}}$ | $LN_{x_{min}-PL}$ |
| Minimum | 0.47 | $-0.52$ | $-519.53$ | 0.88 | 0.80 | 1.07 |
| First Quartile | 0.48 | 0.48 | $-230.57$ | 0.93 | 0.87 | 1.46 |
| Median | 0.51 | 0.52 | $-6.49$ | 0.95 | 0.93 | 1.96 |
| Third Quartile | 0.53 | 0.71 | $-2.65$ | 0.94 | 0.96 | 9.27 |
| Maximum | 0.57 | 0.84 | $-0.02$ | 0.99 | 1.16 | 14.02 |
| Mean | 0.51 | 0.53 | $-133.67$ | 0.94 | 0.92 | 4.99 |
| Stdev | 0.03 | 0.27 | 208.53 | 0.03 | 0.08 | 5.11 |
| Skewness | 0.31 | $-2.04$ | $-1.05$ | $-0.28$ | 0.64 | 0.89 |
| Kurtosis | $-1.09$ | 5.67 | $-0.80$ | $-0.31$ | 0.85 | $-1.06$ |
| *p*-value (Shapiro-Wilk) | 0.045 | $2.2 \times 10^{-5}$ | $7.63 \times 10^{-7}$ | 0.43 | 0.09 | $6.2 \times 10^{-6}$ |

Figure 3 shows a visualization of the log–log plots of the CCDF ($P(d_v \geq x)$) of the three cases at $G_1$ and $G_{30}$. In Case 1, the PL model, when it is fitted from $x_{min} = 1$, is a bad fit for the data; in Case 2, the range of LN fitting is greater than the range of PL fitting; in Case 3, there is no significance difference between the two models.
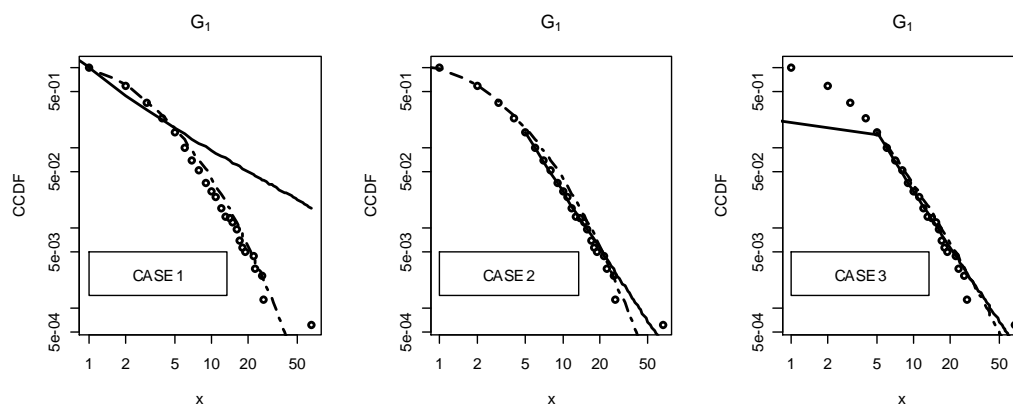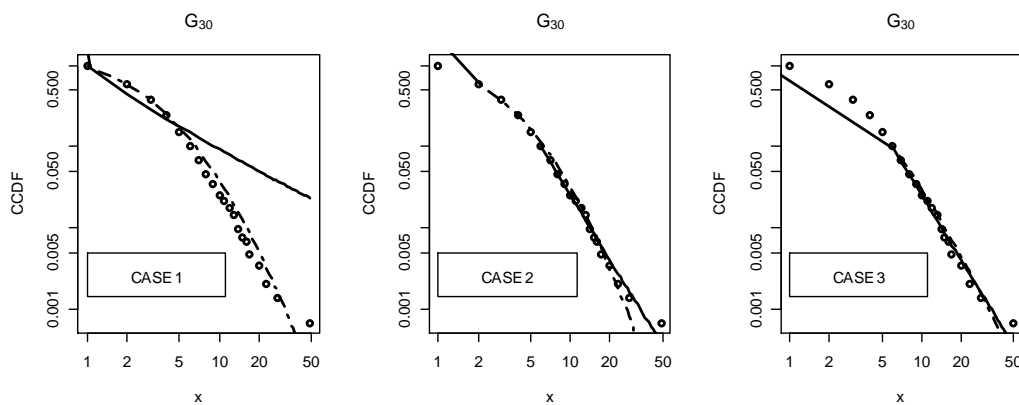


**Figure 3.** *Cont.*

**Figure 3.** Visualization of the log–log plots of the complementary cumulative distribution function (CCDF). The dashed lines refer to the CCDF distributions of the LN model and the solid line refers to the PL model.

## 4. Discussion

The best way to summarize the distinct number of calling partners per client per day, when considering the evolution of the network graph in time, is via an LN model, even though it was rejected 2 more times than the PL model was. This is based on the evidence that *p*-values, which are used to define the LN model as plausible or not, were always reliable. In the PL model fittings, six times out of 30, the *p*-values were not reliable. Furthermore, the range of data that is modeled via LN was always greater than that modeled via PL; furthermore, when the tails of each distribution were compared, no significance differences were found. Therefore, it cannot be said that the distributions of the LN parameters are normal. It would be interesting to define a distribution that models the parameters of LN as the best-fitted degree sequence distribution of the network graphs series.

**Supplementary Materials:** Data records used for this study are available at DOI: 10.13140/RG.2.2.29159.55208/1.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1.　Nanavati, A.-A.; Singh, R.; Chakraborty, D.; Dasgupta, K.; Mukherjea, S.; Das, G.; Gurumurthy, S.; Joshi, A. Analyzing the Structure and Evolution of Massive Telecom Graphs. *IEEE Trans. Knowl. Data Eng.* **2008**, *20*, 703–718. [CrossRef]
2.　Nanavati, A.-A.; Gurumurthy, G.-D.; Das, G.; Chakraborty, D.; Dasgupta, K.; Mukherjea, S.; Joshi, A. On the Structural Properties of Massive Telecom Call Graphs: Finding and Implications. In Proceedings of the 15th ACM International Conference on Information and Knowledge Management, CIKM'06, Arlington, VA, USA, 6–11 November 2006.
3.　Seshadri, M.; Machiraju, S.; Sridharan, A.; Bolot, J.; Faloutsos, Ch.; Leskovec, J. Mobile Call Graphs: Beyond Power-Law and Lognormal Distributions. In Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'08, Las Vegas, NV, USA, 24–27 August 2008.
4.　Dong, Z.-B.; Song, G.-J.; Xie, K.-Q.; Wang, J.-Y. An Experimental Study of Large-Scale Mobile Social Network. In Proceedings of the 18th International Conference on World Wide Web, WWW 2009, Madrid, Spain, 20–24 April 2009.
5.　Noka (Jani), E.; Hoxha, F. Comparative Analysis of the Structural and Weighted Properties in Albanian Social Networks. *J. Multidiscip. Eng. Sci. Technol.* **2016**, *3*, 4505–4509.
6.　Onnela, J.-P.; Saramäki, J.; Hyvöven, J.; Szabó, G.; Lazer, D.; Kaski, K.; Kertész, J.; Barabási, A.-L. Structure and Tie Strengths in Mobile Communication Networks. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 7332–7336. [CrossRef] [PubMed]

7. Onnela, J.-P.; Saramäki, J.; Hyvönen, J.; Szabó, G.; Argollo de Mendez, M.; Kaski, K.; Barabási, A.-L.; Kertész, J. Analysis of a large-scale weighted network of one-to-one human communication. *New J. Phys.* **2007**, *9*, 179. [CrossRef]

8. Aiello, W.; Chung, F.; Lu, L. A random graph model for massive graphs. In Proceedings of the 32nd Annual ACM Symposium on Theory of Computing, New York, NY, USA, 21–23 May 2000.

9. Aiello, W.; Chung, F.; Lu, L. A random graph model for power law graphs. *Exp. Math.* **2001**, *10*, 53–66. [CrossRef]

10. Gjermëni, O.; Ramosaço, M.; Zotaj, D. Power-Law versus Lognormal Distribution in a Phone Call Network Graph. In Proceedings of the International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE), Sofia, Bulgaria, 13–14 November 2015.

11. Cortes, C.; Pregibon, D.; Volinsky, C. Communities of Interest. In *Advances in Intelligent Data Analysis*; Springer: Berlin, Germany, 2001; pp. 105–114.

12. Ye, Q.; Zhu, T.; Hu, D.; Wu, B.; Du, N.; Wang, B. Cell Phone Mini Challenge Award: Social Network Accuracy—Exploring Temporal Communication in Mobile Call Graphs. In Proceedings of the IEEE Symposium on Visual Analytics Science and Technology, Columbus, OH, USA, 19–24 October 2008.

13. Newman, M.-E.-J. The Structure and Function of Complex Networks. *SIAM Rev.* **2003**, *45*, 167–256. [CrossRef]

14. Royston, P. An extension of Shapiro and Wilk's W test for normality to large samples. *Appl. Stat.* **1982**, *31*, 115–124. [CrossRef]

15. Royston, P. Algorithm AS 181: The W test for Normality. *Appl. Stat.* **1982**, *31*, 176–180. [CrossRef]

16. Royston, P. Remark AS R94: A remark on Algorithm AS 181: The W test for normality. *Appl. Stat.* **1995**, *44*, 547–551. [CrossRef]

17. Bulmer, M.-G. *Principles of Statistics*; Dover Publications: New York, NY, USA, 1979.

18. Balanda, K.-P.; MacGillivray, H.-L. Kurtosis: A Critical Review. *Am. Stat.* **1988**, *42*, 111–119.

19. Hanusz, Z.; Tarasińska, J. Impact of Alternative Distributions on Quantile–Quantile Normality Plot. *Colloq. Biom.* **2015**, *45*, 67–78.

20. Foss, S.; Korshunov, D.; Zachary, S. *An Introduction to Heavy-Tailed and Subexponential Distributions*; Springer Science+Business Media: New York, NY, USA, 2013.

21. Clauset, A.; Shalizi, C.-R.; Newman, M.-E.-J. Power-Law Distributions in Empirical Data. *SIAM Rev.* **2009**, *51*, 661–703. [CrossRef]

22. Clauset, A.; Shalizi, C.-R.; Newman, M.-E.-J. Power-Law Distribution in Empirical Data. Available online: http://tuvalu.santafe.edu/~aaronc/powerlaws/ (accessed on 7 June 2007).

23. Shim, J. Toward a more nuanced understanding of long-tail. *J. Bus. Ventur. Insights* **2016**, *6*, 21–27. [CrossRef]

24. Vuong, Q.-H. Likelihood ratio tests for model selection and non-nested hypothesis. *Econometrica* **1989**, *57*, 307–333. [CrossRef]

25. Rmetrics Core Team. R: A Language and Environment for Statistical Computing. Available online: https://www.R-project.org/ (accessed on 21 April 2017).

26. Gillespie, C.-S. Fitting Heavy Tailed Distributions: The powerRlaw Package. *J. Stat. Softw.* **2015**, *64*, 1–16. [CrossRef]

27. Rmetrics Core Team; Wuertz, D.; Setz, T.; Chalabi, Y. fBasics: Rmetrics-Markets and Basic Statistics; R package 3011.87. Available online: https://CRAN.R-project.org/package=fBasics (accessed on 29 October 2014).

28. Csardi, G. Igraphdata: A Collection of Network Data Sets for the 'igraph' Package. Available online: https://CRAN.R-project.org/package=igraphdata (accessed on 13 July 2015).

29. Csardi, G.; Nepusz, T. The igraph software package for complex network research. *InterJournal* **2006**, *1695*, 1–9.