BABEŞ-BOLYAI UNIVERSITY

DOCTORAL THESIS SUMMARY

# Statistical universalities in socio-economic systems

*Author:*
GERE István

*Supervisor:*
Prof. Dr. NÉDA Zoltán

*The summary of
a thesis submitted
in fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the
Domain of Physics*

Babeş-Bolyai University
Cluj Napoca, Romania
February 2022

# Statistical universalities in socio-economic systems


## ABSTRACT

Data mining and analysis of some socio-economic processes allowed the observation of statistical universalities and a theoretical description by models borrowed from statistical physics. Income and wealth distribution in modern and past societies were investigated and modeled. Interesting scaling laws for the Internet were also revealed and modeled.

Using exhaustive data for multiple years regarding the income in Cluj county we observed a scaling probability density function. By following the dynamics of the personal incomes an evolutionary model, the Local Growth and Global Reset model (LGGR) was proposed for a realistic description of the observed statistics. The universality of the observed probability density function was checked for other countries as well.

A similar approach was applied to describe wealth distribution in modern societies. Our method allowed a unified description of the probability density function for all wealth categories, and also a successful description of the negative wealth (debt) regime. A consistent modeling of wealth distribution for many countries with different economies were given.

For completing our studies on wealth distribution agricultural and taxation records were analyzed in order to estimate the distribution of wealth in a well defined geographic area, Sâncraiu commune (Cluj county, Romania). Different periods of the commune's economic history was analyzed and modeled. We observed a Gaussian like distribution of wealth in the communist period, and a typical heavy-tailed wealth distribution for the year 2021, 32 years after the abolishment of the communist regime in Romania. The LGGR approach allowed again a good description for the distribution of wealth in each period, with kernel functions for the growth and reset rates that are consistent with the economic policies of the investigated periods. Socio-economic inequality measurements were investigated and compared successfully with the model results.

Finally, the last study presented in the thesis deals with the evolution of the Internet's wiring topology. We uncovered interesting scaling laws and offered a simple and economically realistic model, that produces graphs with similar properties to the Internet on a router level. This model is capable of reproducing the degree distribution observed in simple experiments along with the scaling that is observed for the number of intermediary router points as a function of the physical distance between the routers in the wired Internet.

This thesis confirms again the usability of simple statistical physics models in understanding universalities in complex socio-economic phenomena.

Keywords: *statistical univestalities, econophysics, sociophysics, complex systems, statistical physics, distribution of income, distribution of wealth, local growth and global reset model, wealth estimation methods, scaling laws, networks, Internet, agent-based modelling*

# Contents of the Thesis

# Contents of the summary

# Introduction

Abundant data sources available for us and handled by today's powerful computers allow us to reveal statistical universalities in many interdisciplinary areas of study [1]. These statistical observations can be analyzed and understood with the methods of statistical physics.

The most basic of such a statistical universality even pre-dates the computerised age. The famous **Gaussian distribution** is observed in many unrelated phenomena, from the distribution of the chest girth of soldiers [2], to the distribution of elapsed times between cyber attacks against our Departments file-server. The Central Limit Theorem states that if a random variable is the sum of multiple independent random variables it's distribution will be a Gaussian one [3]. From this point of view the universal nature of the Gaussian distribution is clear, phenomena influenced by random underlying factors follow such a distribution.

Another example of generally known statistical universities is the **Exponential distribution**. It's universality lies in the fact that it is characteristic to many processes that are memoryless. As such it is a prime way to describe waiting times for events that happen with a constant rate, like elapsed time for a device's component to fail [3], or distribution of elapsed time for a single radioactive atoms to decay [4].

At the center of the work presented in this thesis lies the **power-law**. Many complex and unrelated phenomena [11] produces distributions that converges towards a power-law for great values.

$$\rho_{PL}(x) \propto x^{-a} \tag{1.1}$$

Notable examples of phenomena presenting power-law like distributions (Eq. 1.1.) are the citation distribution of scientific articles, the "share" distribution of Facebook posts [5], the distribution of energy released by earthquakes (Gutenberg and Richter law) [6], the distribution of city population sizes [7], travel speeds on transportation networks [8] or the famous observation of Pareto the distribution of the wealth [9].

As pure power-law like distributions cannot be normalised on the interval $[0, \infty)$ the experimentally observed distributions are usually described with a functions that converge towards a power-law. Examples for such functions are: the Beta-Prime distribution, Tsallis-Pareto Type 2. (Lomax) distribution to name a few.

Heavy-tailed distributions are detected by careful analysis of data relating to real world phenomena. We usually use the histogram method, with different binning methods. Alternatively one can construct the cumulative probability density distributions as these also present a power-law like shape. Pareto originally used cumulative curves to describe the wealth distribution in his studies [9].

The framework for studying power-laws are found in statistical physics. Power-law functions are usually characteristic for critical phenomena that are around phase transitions, when a relevant quantity of the system diverges as a function of another quantity [10].

It seems from the observations that power-laws are embedded into the complex systems. This is understood through the concept of Self Organized Criticality (SOC), a statistically stable dynamic state towards which many systems tend [11]. In such view many socio-economic systems can be viewed as **complex systems** [12].

A challenging possibility for collecting experimental data is that such data are electronically available regarding social and economic systems, by the state and private institutions. This allows for data mining that is necessary to reveal statistical properties, and the underlying processes guiding these systems. In this work we intend to further investigate the generality of the power-law type distributions in socio-economic systems, considering specific problems and models that could further confirm their universality.

In the second and third chapters of this thesis we describe studies on wealth and income distribution in well-delimited human societies [13,14], describe a novel wealth estimation methodology, valid for small homogeneous communities, and explain the experimentally observed distributions using a recently proposed master equation based model [15].

The fourth chapter is related to the scaling properties of the Internet as a physical network [16]. We aim to explain by a simple wiring model some intriguing scaling laws observed experimentally.

# Statistical universalities for income and wealth distributions

Wealth and income are in the constant focus of many socio-economic studies as their distribution affects every scale of our life, from the level of individuals to the society as a whole.

It is known, ever since the famous discovery of Vilfredo Pareto, at the end of the 19th century [9], that the cumulative distributions (Eq. 2.1.) of wealth and income are heavy-tailed ones.

$$P(X > x) \sim x^{-a} \tag{2.1}$$

The power-law like distribution was observed through wealth proxies in historical data referring to building areas in ancient Egypt [17] or in the number of serfs owned by nobles in medieval Hungary [18].

For income we refer to the money received for work, or extracted from investments [19], while for wealth to the quantity of valuables a person posses [20]. The measures of both income and wealth are described as similar heavy-tailed distribution. It also must be noted that wealth and income are not as strongly related as one would expect [21].

**Data mining**

Our experimental studies of income imply an exhaustive social security database from 2001 to 2009 in Cluj county, Romania [22] extended with Hungarian tax data, public income datasets from surveys for United States of America [23] and Finland [24] from the statistics yearbook in case of Russia [25] and public tax data for Australia [26] and Japan [27].

The data regarding the wealth was obtained from a single source, the World Inequality Database [28] regarding the wealth in the United States, Russia and France estimated with complex methodology [29].

**General form of wealth and income distributions**

The shape of income and wealth distribution was studied extensively in the last few decades in econo-physics [30]. It is usually considered to be a probability density distribution (p.d.f) with a power-law tail, and an exponential mid-range. We study these as regions in the p.d.f, with an added region for the case of very low or negative values. We present these on Fig. 2.1. Researchers in the field of econo-physics applied the methods of statistical physics for a deeper understanding of this characteristic shape. [30–33].

**Building a proper and realistic model**

In our approach we consider that a good model for wealth or income distribution has to be founded on realistic assumptions. The resulting p.d.f. should converge towards a power-law in case of the large values. It should accommodate negative wealth

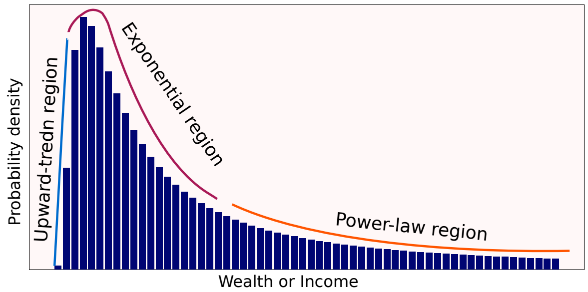


**Figure 2.1:** *A schematic illustration, representing the three regions defined in the wealth and income distribution function.*

values too to explain debt. The distribution should be continuous on the whole interval of the wealth or income values.

## 2.1 The master equation approach, local growth and global reset model

Our studies are based on the master-equation approach [15]. This evolution equation is based on a Local Growth between consecutive states and a Global Reset to a new state, hence the abbreviation of **LGGR**. The equation of evolution for the likelihoods $P_k(t)$ (the probability of a constituent of the system to own $k$-quanta) can be written as:

$$\frac{dP_k(t)}{dt} = \delta_{k,0}\overline{\lambda}(t) - \lambda_k P_k(t) + \nu_{k-1}P_{k-1}(t) - \nu_k P_k(t) \tag{2.2}$$

In Eq. 2.2. the term $\nu_k$ presents the growth rate from state with $k$ owned units to state with $k+1$ units, while $\lambda_k$ represents the reset rate to the state where $k=0$. The process is presented on Fig. 2.2.

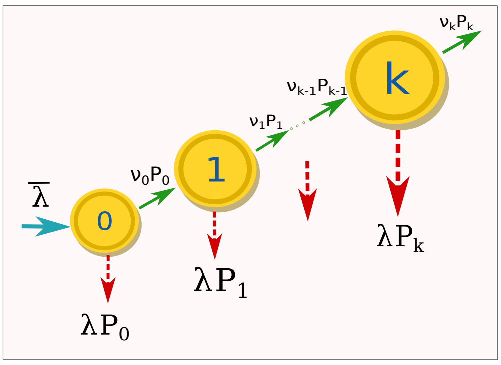


**Figure 2.2:** *Visualisation of the LGGR process between the states described in Eq.2.2.*

We generalized the evolution equation for the case of continuous states, and calculate the form of the stationary p.d.f [15,36]:

$$\rho_s(y) = \rho_s(0)\frac{\nu(0)}{\nu(y)}e^{-\int_0^y \frac{\lambda(z)}{\nu(z)}dz} \tag{2.3}$$

### 2.1.1 Application of the LGGR for income distribution

The exhaustive databases for Cluj county, Romania were used to study the dynamics of the income. The income of Individual employees was followed through the yearly databases to describe the dynamics. By observing the yearly growth of the salaries (Fig.2.3) we concluded for the growth rate $\nu(y)$ a form of a linearly increasing function (Eq.2.4.).

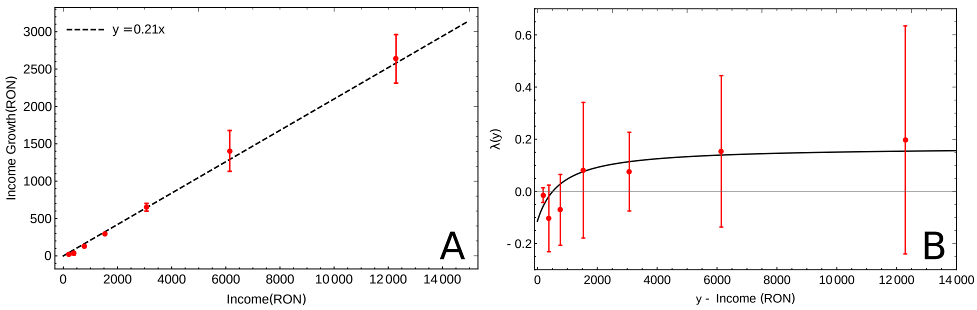


**Figure 2.3:** ***A****: The growth rate as a function of the mean income. On the X-axis we have the average salary of the k-th employee group, in RON and the Y-axis is the average yearly income change for the k-th group ($\Delta_k i$).* ***B****: Plot of the experimentally determined reset rates. On the X-axis we have the k-th employee group as Income(RON), and on the Y-axis we have the average reset rate for the k-th group $\lambda(y)$. The black line represents the fited $\lambda(y)$ function. The red bars are the standard deviation from the mean value.*

$$\nu(y) = \beta(y+h) \tag{2.4}$$

By following the number of incomers and the retirees from the workforce in a given income range we computed experimentally the reset-rate $\lambda(y)$ (Fig.2.3). The reset rate can be approximated in the form given by Eq. 2.5. This "smart" reset-rate is negative for low income while positive for large values.

$$\lambda(y) = S - \frac{a}{y+c} \tag{2.5}$$

With the experimentally observed kernel functions we obtain the form of the p.d.f. for the income values scaled by the mean:

$$\bar{y}\rho_s(y) = (\frac{m-n}{n-1})^{m-n}\frac{\Gamma(m)}{\Gamma(m-n)\Gamma(n)}(1+\frac{y}{\bar{y}}\frac{m-n}{n-1})^{-m}(\frac{y}{\bar{y}})^{m-n-1} \tag{2.6}$$

Rescaling the income distribution data to the mean income of the respective year, the data points collapse on a single curve. The collapsed data can be described with the Eq. 2.6, which is a Beta Prime distribution. For Cluj county the distribution is found in the simple form:

$$\bar{y}\rho_s(y) = 12\frac{y}{\bar{y}}(1+\frac{y}{\bar{y}})^{-5} \tag{2.7}$$

The same experimental distributions were constructed for Hungary and for Japan (Fig. 2.7). With the condition $n = m - 2$ we can fit these with a single $m$ parameter.

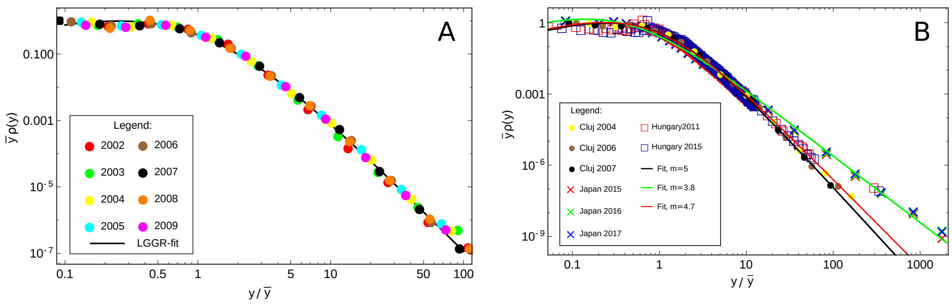


**Figure 2.4:** ***A**: The experimental probability density for the income rescaled with the yearly mean, for Cluj county. The fit is the black line, and is given by Eq. 2.7. **B**:Yearly experimental income distributions rescaled with the yearly mean, for Cluj county, Japan and Hungary. For Cluj county $m = 5$, for Japan $m = 3.7$ and for Hungary $m = 4.7$.*

## 2.1.2 Application to wealth distribution

Similarly with income we normalized the wealth values with the yearly mean $r = \frac{R}{\bar{R}}$. The resulting distributions collapse. An example is presented for USA in Fig. 2.5.

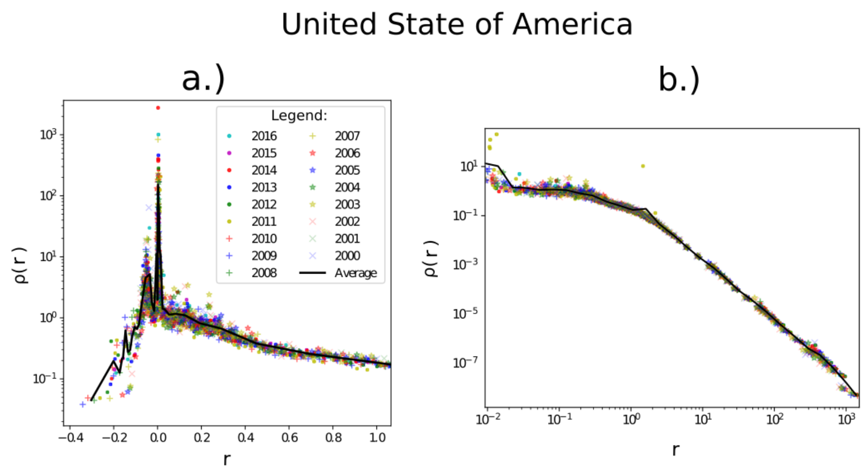


**Figure 2.5:** *The rescaled probability density function for wealth distribution $r = \frac{R}{\bar{R}}$ in the case of USA. The averaged p.d.f. for the presented years is denoted as a black line. **a.)** The lower end of the distribution (below the average), the negative wealth region (debt) can be presented using a log-normal scale. **b.)** The probability density function for the positive part of wealth on a log-log scale.*

### Linear growth rate for the wealth dynamics

The growth process was studied for 22 wealthy individual from the yearly compiled Forbes billionaires list [34] for the years 2001-2019. A linear preferential growth, was observed, similar to the one in the case of income (Eq. 2.4). Intuitively we assume the Reset rate in the same form here as the one used in modelling income distribution (Eq.2.5)

Considering the condition $c = h$ in Eq. 2.4 and Eq. 2.5 allows wealth to grow on the interval $[-h, \infty)$, and the reset can introduce individuals with negative wealth. We enter these rates into Eq. 2.3, and calculate the stationary p.d.f of wealth as Eq. 2.8.

$$\rho_s(y) = \frac{a^S}{\Gamma(S)} e^{-\frac{a}{h+y}} (h+y)^{-1-S} \tag{2.8}$$

In Eq. 2.8 we rescale with the mean wealth, and use the notation $r = \frac{y}{\bar{y}}$ for the wealth variable. With the further notations $\kappa = \frac{a}{h(S-1)} - 1$ and $\bar{y} = \kappa h$, the p.d.f of the wealth scaled with the mean is found in the form of Eq. 2.9.

$$\rho(r) = \frac{\kappa(\kappa+1)^S (S-1)^S}{\Gamma(S)} e^{-(\kappa+1)\frac{S-1}{1+\kappa r}} (1+\kappa r)^{-1-S} \tag{2.9}$$

The scaling with the mean changes the domain on which the wealth distribution function is defined:

$$[-h, \infty) \to [-\frac{1}{\kappa}, \infty) \tag{2.10}$$

### Comparison with the experimental results

We use the previously established p.d.f (Eq. 2.9) to describe the experimentally obtained, averaged and rescaled wealth. We present averaged trends for wealth distribution in the United States of America, and Russia together on Fig. 2.6. The best fitting parameters are $S = 1.4$ and $\kappa = 6.5$. The fit qualitatively describes the negative wealth region along with the heavy-tail.

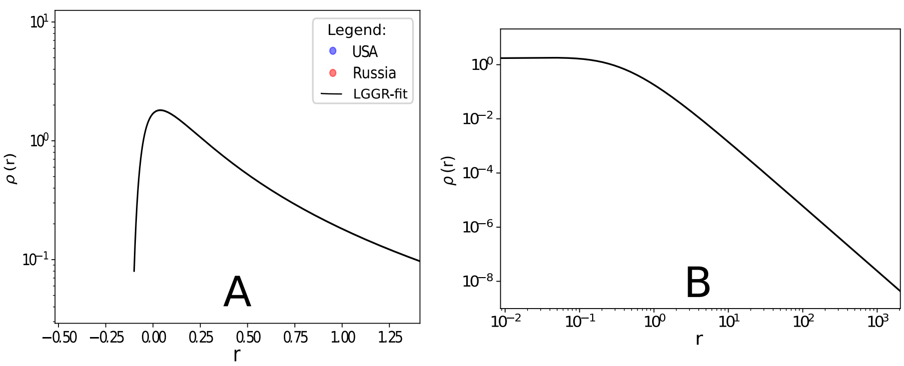


**Figure 2.6:** *The probability density function for the distribution of the rescaled wealth for USA and Russia. The p.d.f.s averaged for different years are plotted with the symbols. The trends for USA and Russia are rather similar and can be described by our LGGR-model, by choosing the parameters $S = 1.4$ and $\kappa = 6.5$ in Eq. 2.9.* ***A****: Region of low and negative income is presented on log-normal scale.* ***B****: Region of positive income on a log-log scale.*

Our model can be compared with the famous Mezard & Bouchaud [35] (MB) approach (based on wealth exchange between all agents and a multiplicative random growth process). The MB approach offers a one parameter fit:

$$\rho(r) = \frac{k^k}{\Gamma(k)} e^{-\frac{k}{r}} r^{-2-k} \tag{2.11}$$

On Fig. 2.7 the LGGR and MB model is compared with experimental data. Both models

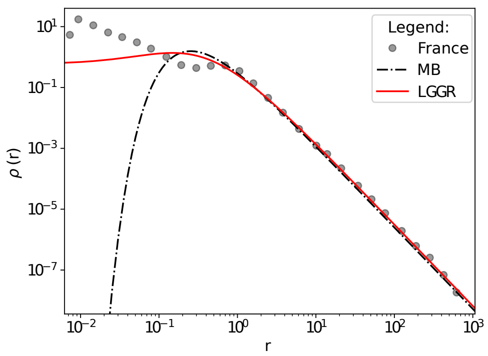


**Figure 2.7:** *The application of the LGGR and the MB model for the case of France. The fitting parameters in this case are $\kappa = 7, S = 1.68$ for the LGGR, and $k = 0.86$ for the MB model.*

produce the same power-law tail for the distribution, with a similar exponent: $-2.68$, therefore both of these methods describe the rich part of the distribution well. The MB approach is only valid on $[0, \infty)$. The p.d.f. obtained from the LGGR approach in this case implies the existance of a negligible amount of negative wealth $F_{neg} \propto 0.063$ for $\kappa = 7$ and $S = 1.68$, and describes better the lower region of the wealth distribution.

## 2.2 Conclusion and discussion

Experiments and theoretical modeling revealed interesting stylised facts for income and wealth. By rescaling income and wealth with their mean values, intriguing universalities are revealed in the distribution function. A remarkable similarity is observed in the wealth distributions for United States and Russia, which might well be a pure statistical coincidence.

The exhaustive data allowed us to observe the guiding principles of income dynamics. We described a part of the observed dynamics with a linearly increasing growth rate (also called the Matthew principle formulated as "rich gets richer"). The reseting dynamics observable in the system is well described with a smartly chosen reset rate that is supported by the data.

These kernel functions for the LGGR modelling approach allowed to construct a realistic p.d.f. for income. The same rates were used in the case of wealth, after modifying the parameters to allow for negative values in the distribution. The simple and intuitive assumptions used in our model produced qualitatively good fits for the rescaled income and wealth distributions.

The shortcoming of the LGGR model is that it does not account for inflation or economic growth, a phenomenon which is observed in the experimental data.

# A case study for wealth estimation and modelling

It is likely that biases are introduced while the wealth is estimated with the methods used to create the WID data [37], used for comparison previously. As a consequence some of our results presented in the previous section might also be influenced by these biases.

To our knowledge the literature lacks long term historical studies built upon exhaustive wealth datasets. One may also be interested in the wealth distribution of economic systems, other than the free market economy. In our studies we aimed to fill in this yet uncharted territory of econophysics.

## Exhaustive wealth data from agricultural records and taxation databases

We constructed an exhaustive and realistic wealth dataset, spanning different economic conditions. The local authorities of Sâncraiu (a commune from Cluj county, Romania) allowed us to access the agricultural surveys of 1961 and 1989 , along with the taxation database of the year 2021 in an anonymised manner. The digitalistation of the agricultural records has been done by our team.

Commune Sâncraiu consisits of five villages and is found in a hilly basin South of the town Huedin. An administrative map of the commune is presented in Fig. 3.1.

Although, the land is suitable for agriculture only to a limited extent, historically the locals led a prominently agricultural lifestyle. In such a society, it can be assumed that the bulk of the wealth consists of the lands and animals along with the houses, barns and auxiliary buildings that are owned.

## 3.1 Wealth estimation methodology

The agricultural records for 1961 and 1989 contain the agricultural valuables owned by each household of the commune by categories. These categories for example are: the size of land owned, of house owned, of auxiliary buildings, number of livestock by type, etc. To estimate the wealth owned by a single household $W$ we used the weighted $w_i$ sum of the valuable categories $C_i$ from the agricultural records:

$$W = \sum C_i w_i \tag{3.1}$$

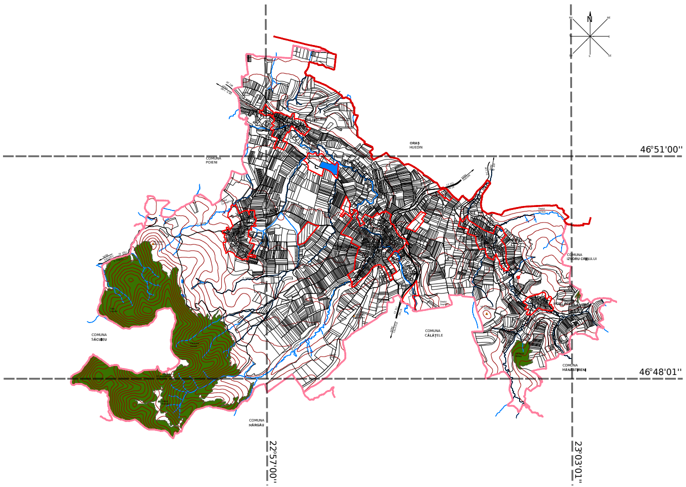


**Figure 3.1:** *Administrative map of Sâncraiu commune, presenting the land parcels, and village territory. Provided by the local authorities.*

The weighting factors describe the relative value of the different wealth categories. By the summation of the weighted valuable categories a unified wealth measure is created, which may be considered as a kind of "currency". We will refer to this unified wealth measure, or "currency" as "wealth units" *(w.u.)*. The valuable categories documented in the agricultural records do change through the years, and as a direct consequence their corresponding value expressed in *w.u.* changes as well.

Due to the agricultural nature of the area, in the years 1961 and 1989 we might consider that the agricultural records documents realistically the bulk of the household wealth. For the year 2021 we use anonimised taxation records to estimate the individual wealth. It has to be noted that the agricultural and taxation records do not contain information regarding the debt (negative wealth) owned by individuals. The wealth is estimated with the same methodology for 2021, but with different valuable categories (value of the buildings, and land areas owned by a taxpayer).

A chosen weight parameter set may overestimate the value of certain valuable categories, leaving the other categories underestimated. A realistic wealth estimation is done by considering multiple different and realistic weight parameter sets. For each year 10 different weight sets were considered, that allowed to construct error-bars in our wealth distribution.

The share of different valuable categories in the total wealth is strongly influenced by the value of the weight parameters. Different composition of the total wealth allows to account for the influence of the errors in the estimation of the wealth of the individuals. In the thesis the effect of the parameter choices are detailed and summarized in some dedicated tables.

The individual household wealths were calculated, with each weight set for the years 1961, 1989, 2021. By using the histogram method, with a fixed number of bins $n_b = 30$, we calculate the distribution of the differently weighted household wealth $\rho(W)$. The effect of the different weight sets causes the middle of the bins to shift along the x-axis as well, resulting in the determination of error bars in both directions, not only on the y-axis. The range of deviations from the mean values are presented as bars placed on the averaged value. This technique allows to present simultaneously the spread in the different wealth distributions originating from the different weight sets. We show these results in Fig. 3.2. This method of plotting reveals a clear trend with

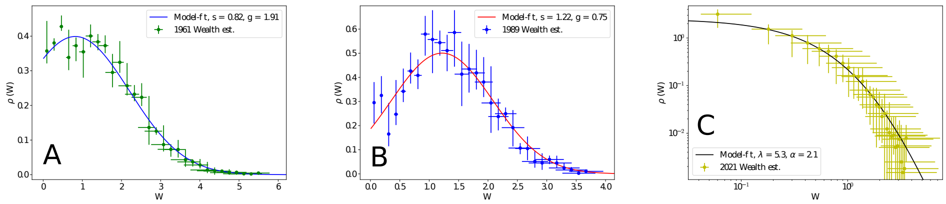


**Figure 3.2:** *Normalized probability density function of the estimated wealth for year 1961 (**A**), 1989 (**B**), and 2021 on a log-log plot (**C**). The bars on the points represent the uncertainty rising from the different weight sets. The continuous line is the fit obtained by our model.*

the relevant uncertainty. The shape is not affected significantly by the different weight sets, although the exact location of the datapoints do shift. We will come back to this observation in a later section.

## 3.2 Theoretical approach for the observed distributions

The LGGR modelling framework can provide a theoretical description for the underlying dynamics of wealth distribution in the different economic conditions.

#### Private wealth under the communist regime? - Constant growth rate and a linear reset rate

The socio-economic effects of the communist policy can be accounted with proper growth and reset rates. The wealth of the household grows slowly. This growth is independent of the wealth of the household and can be described with constant growth rate $\nu(x) = g$. It describes a controlled growth in wealth for each element of the society.

The assumption that every member of the society should have roughly close amount of wealth can be also accounted with the reset rate as it controls the overgrowth of the privately owned wealth. We can define an ideal amount of private wealth, that should be enforced to regulate the wealth for the whole society. This is described with a simple linear function $\lambda(x) = x - s$. With these rates the stationary p.d.f. of the wealth in this case is given as:

$$\rho_{s_{com}}(x) = \sqrt{\frac{2}{g\pi}}[1 + \text{erf}(\frac{s}{\sqrt{2g}})]^{-1} e^{\frac{-x(x-2s)}{2g}} \tag{3.2}$$

We present the fit $\rho_{s_{comm}}$ along with the experimental results on Fig. 3.2. for 1961 and 1989. The fit parameters together with the experimentally obtained means, and the theoretical ones are listed in Table. 3.1. It is interesting to note that the value of the

| Year. | $g$ | $s$ | $R^2$ | $\overline{x}_{theo}$ | $\overline{x}_{exp}$ |
|---|---|---|---|---|---|
| 1961 | 1.91 | 0.82 | 0.98 | 1.459 w.u. | [1.36; 1.49] w.u. |
| 1989 | 0.75 | 1.22 | 0.93 | 1.359 w.u. | [1.199; 1.391] w.u. |

**Table 3.1:** *The best fit parameters from Eq.3.2 for the years 1961 and 1989, and the resulting average wealth from theory $\overline{x}_{theo}$, and the interval of $\overline{x}_{exp}$ suggested by the experimental data with various weight-parameter sets.*

average wealth and the value of the reset parameter - $s$ are quite similar, especially in the case of 1989.

**The rich gets richer dynamics - Linearly increasing growth rate and a constant reset rate.**

For modelling the dynamics of wealth as a stationary situation in 2021, we re-consider the approach used in Chapter. 2 for the free-market economy in modern societies. In this case however there are no individuals with debt (negative wealth) in the commune. We chose the linearly increasing growth rate $\nu(x) = x + \alpha$, describing the preferential nature of the economic interactions. A simple constant reset rate is considered $\lambda(x) = \lambda$. For the studied small society, this means that resetting has the same probability in every state independently of the owned wealth. We obtain a compact solution of the dynamical evolution in the stationary limit [15]:

$$\rho_{s_{cap}} = \frac{\lambda}{\alpha}(1 + \frac{x}{\alpha})^{-1-\lambda} \tag{3.3}$$

This particular distribution is the well known Tsallis-Pareto (also known as Lomax type 2. distribution) which is the generally expected distribution in systems with preferential growth dynamics. We present the fit $\rho_{s_{cap}}$ in comparison with the experimental results on Fig.3.2. The fit parameters along with the experimentally obtained mean, and the theoretical one are presented in Table. 3.2.

| Year. | $\lambda$ | $\alpha$ | $R^2$ | $\overline{x}_{theo}$ | $\overline{x}_{exp}$ |
|---|---|---|---|---|---|
| 2021 | 5.3 | 2.1 | 0.91 | 0.488 w.u. | [0.297; 0.682] w.u. |

**Table 3.2:** *The parameters chosen for the Tsallis-Pareto fit from Eq. 3.3. for the data from 2021, and the resulting average wealth from theory $\overline{x}_{theo}$, and the interval of $\overline{x}_{exp}$ given by the different weight parameter sets.*

Sample sizes: in 1961 a total of 1133 individual households, in 1989 some 921 households and the tax records of 2021 counted 1595 individual taxpayers.

## 3.3 Observed inequalities

We calculated the value of the most well known inequality measures based on the experimental and theoretical results. In Table 3.3 we summarise the Gini coefficient $G$ [38], Pareto-point $P_p$ (the $P_p$ share of society that owns the $1 - P_p$ share of wealth) comparing the experimental values and the ones offered by our model. The inequality visualising tool, the Lorenz-curve (the cumulative fraction of wealth $F(x)$ as a function of the cumulative share of society that owns it $C(x)$) [39] is presented in Fig. 3.3.

| Year | $G$ | | $P_p$ | |
|---|---|---|---|---|
| | Theoretical | Experimental | Theoretical | Experimental |
| 1961 | 0.378 | [0.377;0.379] | 0.367 | [0.366;0.368] |
| 1989 | 0.312 | [0.304;0.315] | 0.391 | [0.390;0.395] |
| 2021 | 0.552 | [0.543;0.579] | 0.298 | [0.282;0.299] |

**Table 3.3:** *The Gini coefficient and Pareto point for the years 1961, 1989 and 2021. The experimental values are calculated from the wealth estimations, while the theoretical values are obtained from the theoretical fits of the LGGR predictions.*

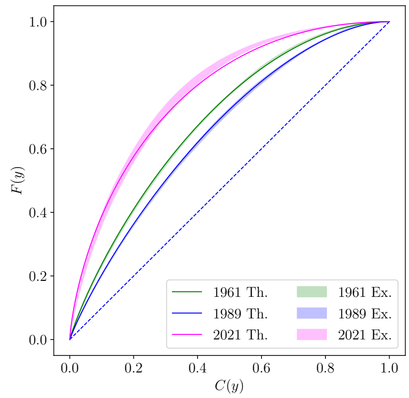


**Figure 3.3:** *The Lorenz curves for 1961, 1989 and 2021, from the experimentally estimated wealth are indicated by a shaded region (the spread produced by the weight parameters). Our theoretical estimate obtained with the LGGR model is indicated by the continuous line.*

The inequality metrics show a decrease in wealth inequality form 1961 to 1989. As it is expected, the free market economy led to an increased inequality by the year 2021. These changes are reflected through the resulting p.d.f.s from the LGGR model.

## 3.4 Discussion and conclusion

The observed wealth distributions are representative for three different socio-economic periods of Romania. In the year 1961 the communist government enforced full collectivization, the existing wealth structures in this year are the result of the Land reforms of 1921 and 1945. The year 1989 is the last year of the communist regime in Romania, where collectivization largely reduced the private wealth. The last studied year 2021, is 32 years after the end of communism and after the transition to a capitalistic free-market economy. For the years 1961 and 1989 the experimentally estimated p.d.f. of wealth is quite different from the one we observed in modern societies, and follows a Gaussian-like shape. For year 2021, the wealth distribution has the characteristic Pareto tail.

The use of different weight sets was crucial for the estimation of the total wealth, as we have no information regarding the exact market values of the valuable categories. This method of superimposing multiple estimations allowed us to draw qualitative conclusions regarding the shape of the wealth distribution. Remarkably it did not alter the qualitative trends of the distribution. This stability of the observed shape of the p.d.f. might be the consequence of the fact that the economy of 1961 and in part 1989 was based on the agriculture. There is a direct relation between the existing valuables, since land area determines the amount of other owned goods as well. In the years befor 1989 free trade of land was non-existent and most of household income was invested in the most valuable and still private object, the house. This is observed in the data as the total private house area in 1961 was $44273m^2$ which grew to $65396m^2$ in 1989. By 2021 the local economy is not dominated by agriculture anymore, new sources of income, that creates wealth, appeared such as tourism, and services. This introduces a greater spread in the experimental mean wealth, resulting in the widening of the error-bars on the plot. This is probably a result of the reduced number of categories in the taxation record, and the diversification of the local economy. In simplified terms the valuable categories became de-coupled from the value of the land.

### The LGGR approximation and the justification of the rates

By considering proper kernel functions in the growth and reset rate for the LGGR model we reproduced successfully the p.d.f. of wealth for both 1961 and 1989. The dynamics achieved by the considered characteristic processes is a simplified picture for the economic situation under the communist regime. We observe the decreasing of the inequalities from 1961 up to 1989 as a decrease of the parameters $g$ and the increasing reset rate $s$ (see Table 3.1). The distribution function peaks at $s$ desired wealth value, the greater the value of $s$ the maximum of the distribution gets closer to the average value of the wealth, reducing the inequality in the system.

We assume that by 2021 the guiding principle of wealth accumulation is the preferential growth (Matthew principle), described as a linear growth rate (discussed in the previous Chapter in detail). The fact that none of the inhabitants of the village had debt allowed us to choose a simple constant reset rate. This approach resulted in a Tsallis-Pareto distribution. The tail exponent "$a$" for the power-law like trend obtained for the Sâncraiu commune can be compared with the ones obtained in modern societies on a country level. This is done in Table 3.4.

| Wealth Data source | Power-law exponent $a$ |
|---|---|
| Wealth data France (2000-2014) | 2.68 |
| Wealth data USA (2000-2016) and Russia (2000-2015) | 2.4 |
| Sâncraiu commune Tax data (2021) | 6.3 |

**Table 3.4:** *The power-law tail exponents of wealth distribution for USA, France and Russia as discussed in Chapter. 2, compared with the tail exponent of the 2021 wealth distribution obtained from the taxation data in Sâncraiu commune in 2021.*

The tail exponent for the case of Sâncraiu commune is higher (meaning lower inequalities) than the one observed at country levels. The difference may be explained by the relatively small size of the commune and lack of the very rich category.

This chapter considered a novel "bottom-up" approach for wealth estimation from simple proxies. Furthermore, we provided economically justifiable kernel functions for the LGGR model, to describe the wealth distribution under different economic situations. The theoretical results reflect the quantitative experimental observations such as the shape of the p.d.f, the mean wealth and inequality metrics.

We emphasize that the authors of this work, does not take side or subscribe to any political policy or ideology that might be connected to this work. We limit ourselves to simply present the facts as they were observed from the agricultural and taxation records.

# Modelling the wiring topology of the Internet

The Internet is based on a physically wired network that interconnects the computers of our modern world. It has been shown that the number of Internet connections a city has is related to socio-economic factors like the GDP [40].

The aim of this chapter is to present an interesting dynamical scaling law and a simple agent based model that takes into account realistic assumptions to generate a graph similar to the real Internet.

#### The Internet

The Internet is a physically existing infrastructure that interconnects computer networks [41]. At the lowest level [42] we consider that it operates by computers sending packets of data towards another computers through routers connected via cables.

## 4.1 Relevant topological and dynamical properties of the Internet

The Internet as a complex system presents several interesting scaling properties. We have unveiled a novel power law like scaling [16] between the Round-Trip-Time (RTT) of a Ping request [43, 44], and the geographic distance between two computers. The Ping request measures the time to access a remote device on the Internet. A data package is sent towards a remote device, than sent back by it. The sending computer measures the elapsed time in milliseconds. In the study 24700 remote devices were "Pinged" over the Internet, distributed all around the Globe. The results were averaged.

We reconsidered this study by using the **Traceroute** tool as well. The "Traceroute" is similar to "Ping", but includes extra features. As the name reveals it traces a route from the staring point towards the destination, by returning a package from every router along the way [45]. We get an RTT for each step of the route. The IPv4 Routed/24 Topology Dataset provided by CAIDA [46] for the year 2017 was used. The Traceroute allowed creating a graph from a reduced section of the Internet with $V = 43992$ vertices $C = 191005$ connections, with an average degree $\bar{c} = 8.68$. The RTTs were extracted from this dataset.

The Traceroute experiment allows us to study the scaling of the RTT with the number of intermediate routers ("steps").

The RTT scaling as a function of distance ($l$) and as a function of the number of intermediary router "steps" ($S$) are shown on Fig. 4.1.

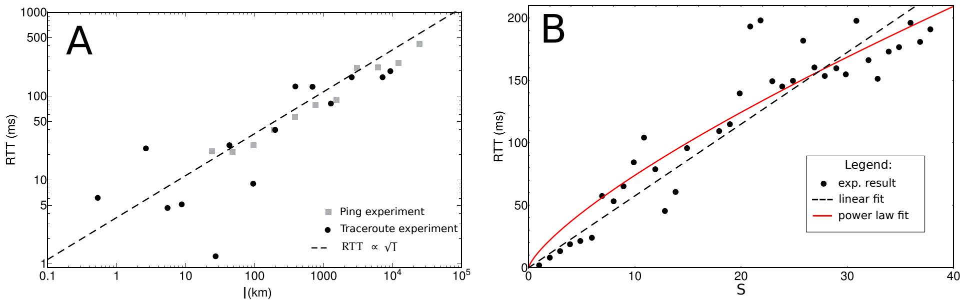


**Figure 4.1:** ***A***:*Averaged RTT (ms) plotted as a function of distance l (km). The results from the Ping and Traceroute methods are presented. The RTT* $\propto \sqrt{l}$ *proportionality holds.* ***B***: *Averaged RTT (ms) as a function of the number of intermediary "steps" on routers S obtained from the Traceroute data. Two fits were considered, the dashed line is a linear fit, the continuous red line is a power-law fit with an exponent* $\frac{3}{4}$. *The coefficient of determination for the power law fit* $R^2 = 0.98$ *is higher than the one obtained for a linear one.*

It is possible to combine the results of the Traceroute experiment with the result of the Ping experiment. These scalings are presented in Fig. 4.1. The observations can be summarised as:

$$
\begin{aligned}
RTT &\propto S^{\frac{3}{4}} \\
RTT &\propto l^{\frac{1}{2}}
\end{aligned}
\tag{4.1}
$$

From Eq. 4.1. one will obtain the scaling of the number of intermediate routers ($S$) as a function of the geographic distance ($l$):

$$
S \propto l^{\frac{2}{3}} \tag{4.2}
$$

## 4.2 Modelling the Internet topology

In order to understand the previously presented observations, we created a simple agent based model for the Internet. In our network based model, the vertices of the graph are the routers attached to the cities while the connections between the vertices represent the cable connections between the cities. Due to the direct proportionality between the GDP of a city and their population, we consider the later as the measure of the cities connectedness to the Internet. In the model the cities are point like and are placed with random coordinates in a simple Eucledian space, with a realistic weight $P_i$ representing its population. The $P_i$ population follows a Tsallis-Pareto distribution with a scaling exponent of $a = -2$. Each city has an imaginary circular area of coverage (maximum extent of possible connections) with a radius proportional with the square-root of the population of the city: $R_i = \sigma\sqrt{P_i}$. We take each possible pair of vertices and calculate a connectivity factor $k_{ij} = \frac{R_i + R_j}{l_{ij}}$, where $l_{ij}$ is the geographical distance between the nodes $i$ and $j$. If $k_{ij} > 1$ a connection is placed between these cities. This graph building process is presented schematically in Fig. 4.2. In our study we considered only the largest connected component of the generated networks *(giant component)*, and studied its properties.

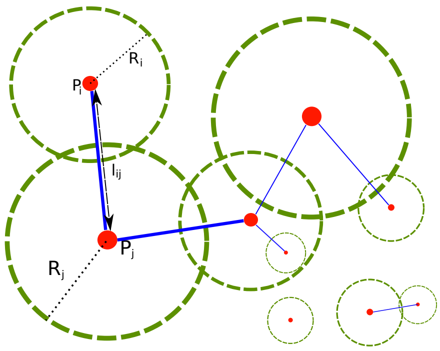


**Figure 4.2:** *Schematic representation of the model. The red dots represent the verticies, the green circles are the area of coverage, and the blue lines are the established connections.*

### Properties of the model graphs

We found that for $\sigma = 0.4$ the average vertex degree of the whole generated system was $\bar{c} \sim 8$, a value very close to the one observed in the experiments. For fixed vertex numbers $M = 2400$ and $M = 8000$ and $\sigma = 0.4$, with an ensamble of 100 graphs we studied than the averaged properties of the generated graphs. The properties of the graphs generated by our model, and those reveled from the Traceroute data fell in between the values found in the literature [47] as shown in Table. 4.1.

| Methods | Properties | | | | |
|---|---|---|---|---|---|
| | Number of Vertices (V) | Number of Connections (C) | Average degree $\bar{c}$ | Maximal degree of a vartex $C_m$ | degree distribution exponent $\gamma$ $f(c) \propto c^{-1-\gamma}$ |
| SKITTER [47] | 9204 | 28959 | 6.29 | 2070 | 1.25 |
| BGP [47] | 17446 | 40805 | 4.68 | 2498 | 1.16 |
| WHOOIS [47] | 7485 | 56949 | 15.22 | 1079 | - |
| Traceroute | 43992 | 191005 | 8.68 | 10474 | 1.23 |
| model, M=2400 | 1946 | 12607 | 12.95 | 1382 | 1.52 |
| model, M=8000 | 6679 | 49748 | 14.88 | 4254 | 1.60 |

**Table 4.1:** *Properties of the Internet graph, with different mapping methods as found in the literature [47]. We compare these results with the properties of the subgraph mapped by the considered Traceroute experiment and the averaged properties obtained by our model.*

The degree distribution $p(c)$ for the generated graphs (averaged from the ensamble) and for a subsection of the Internet (obtained from the Traceroute data) were compared. The Internet is considered to be a scale-free network [47, 48], its characteristic degree distribution is heavy-tailed. A Tsallis-Pareto distribution (or Lomax 2.) distribution offers a good fit, considering the average degree $\bar{c} = 8.68$ obtained from the Traceroute experiment:

$$p(c) = \frac{\gamma}{\bar{c}(\gamma - 1)} (1 + \frac{c}{\bar{c}(\gamma - 1)})^{-\gamma - 1} \tag{4.3}$$

In Fig. 4.3 we plotted the degree distribution of the graph as observed from the Traceroute experiment along with the degree distributions of the graphs created by our model. The model results for different system sizes collapse while the distribution obtained from the Traceroute experiment is slightly different. The graphs generated by our model have thus similar topological properties to the Internet network.

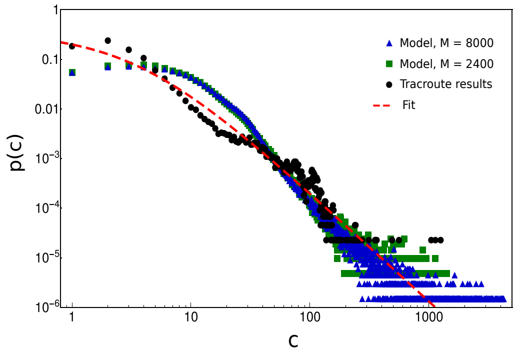


**Figure 4.3:** *The degree distributions in case of the graphs revealed from the Traceroute experiments, and the generated graphs given by our model for $M = 2400$ and $M = 8000$ vertices. The dashed red line is the fitted Tsallis-Pareto distribution with parameters $\bar{c} = 8.68$ and $\gamma = 1.23$. Source: [16]*

## 4.3 Scaling for the information propagation process

The scaling law from Eq. 4.2. can be verified both by the Traceroute experiment, and for the result of our model. The number of visited routers ("hops" or intermediate "steps") as a function of the distance $l$ is shown in Fig. 4.4. In order to compare the results obtained from the model with the results of the Traceroute experiment, the real world distances were rescaled to fit a 1 X 1 square, where simulations have been done.

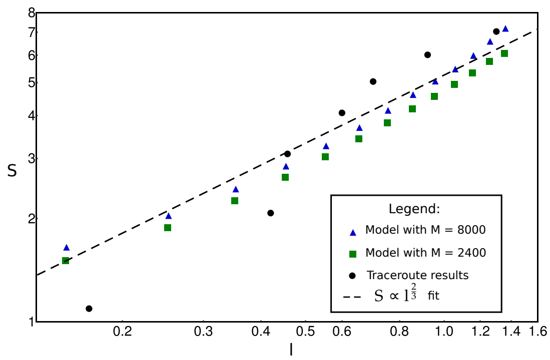


**Figure 4.4:** *Averaged number of "steps" $S$ plotted against the physical distance $l$ between the source and destination, on a log-log scale. The presented results are from the Traceroute experiment where the real 2D geographic space is rescaled on a 1 X 1 square, and from the model where we considered two different number of nodes. The fit reveals the $S \propto l^{\frac{2}{3}}$ scaling. Source: [16]*

Since the shortest path calculation on the graph is analogous to the minimization of the intermediate steps (by the routing policies) during data transfer on the internet, the comparison of the experimental results and the results of the model is feasible.

## 4.4 Conclusion

In this chapter we presented a simple agent based model for the topology of the Internet wiring embedded in the real 2D geographical space. We were capable to generate graphs that capture the essential properties of the Internet topology on a router level.

We also revealed interesting scaling laws between the averaged travel time of information and the geographical distance. Our model successfully reproduced the non trivial scaling, between the number of intermediate routers $S$ and the physical distance (between the source and destination) $l$: $S \propto l^{\frac{2}{3}}$ that was revealed by experiments.

On can also note a few shortcomings of our model. The main problem is that it does not take into account a realistic geography: physical or political boundaries that could have an effect on the wiring topology. Also this approach does not consider weighted connections, meaning that the bandwidths of the real Internet connections are not captured. As a consequence, in the modelled network we cannot identify the high bandwidth Internet backbone, that assures high speed connections at long ranges with low delay [49, 50].

We may conclude however, that this simple approach captures qualitative properties of the Internet, and may be considered as a good first step towards better understanding of the Internet as a network.

# Conclusion

*"PSYCHOHISTORY–...Gaal Dornick, using nonmathematical concepts, has defined psychohistory to be that branch of mathematics which deals with the reactions of human conglomerates to fixed social and economic stimuli.... ... Implicit in all these definitions is the assumption that the human conglomerate being dealt with is sufficiently large for valid statistical treatment." -* **Isaac Asimov [51]**

The above quote described the definition of "psychohistory" a fictional scientific field, the centerpiece of Asimov's famous Foundation trilogy. These words in a certain way encapsulate the contents of our work. It is safe to say that individuals may seem to act in random ways, dictated by some mysterious internal logic. A large number of individuals imply an even larger number of possible choices and interactions, therefore the individuals create a complex system of individuals, called society.

Back than it was in the realm of science-fiction to have such a deep understanding of our society. With the current state of our ever more digitalised world, capturing more and more aspects of our lives, limitless possibilities open up to understand the complexity of our society. These findings stir our curiosity to create models, and explain these phenomena. In the essence of our thinking the individuals are analogous to the particles of a gas, while our society is the bulk to which statistics apply. We draw conclusions regarding our society in the same manner as the Boltzmann distribution describes the energy of the particles in the gas.

Starting from the way of thinking described above at first we looked at the distribution of income for the county of Cluj (Romania), which not surprisingly follows a heavy tailed distribution. The exhaustive nature of the database allowed us to understand the internal dynamics of the income of the workforce. We found a preferential growth in income, simply formulated as the Matthew principle. We understood that employees on average enter the workforce with low income, and exit with larger income. We formulated these findings in a mathematical way, choosing the form of the kernel functions for the growth and reset rates in a simple evolutionary model, the Local Growth and Global Reset model. Using this approach we theoretically describe the stationary probabilty distribution of income. The form of income distributions proved to be appropriate to describe the distribution of income in different countries with different parameters.

By following our way of thinking, from observations on the level of the individuals, we arrived to conclusions on the level of the system that are consistent with the empirical observations regarding the whole system. We observed an interesting universality, namely the fact that by rescaling the income with the mean, for a given region or country, the distributions for different years collapse to a master curve.

We studied the distribution of wealth in modern societies with non-exhaustive data. We observed that the mean rescaled probability distribution of wealth for a given country in multiple years collapsed just like in the case of income. Using the LGGR approach we describe the distribution of wealth. The kernel functions are similar to the ones used in the case of income, but with a slight modification to allow the existence of negative wealth. Our results are comparable to other approaches from the literature.

In our next study we proposed a simple wealth estimation method based on exhaustive records obtained for Sâncraiu commune. It requires that in a given restricted geographical area we know the owned valuables of each household, that are relevant from socio-economic point of view. If the valuables are categorised in pre-defined categories than we can weight these valuables (basically transforming their value into a hypothetical currency) and calculate their linear combination to obtain the wealth of a household. We discovered that such an estimation method produces stable wealth distributions for widely different weight sets (wealth compositions) suggesting a dependence between the valuable categories. This dependence was more articulated for the agricultural society, and became less significant in the modern diversified economy of the same commune. According to our knowledge, this wealth estimation approach is novel.

In this study we discovered a Gaussian-like p.d.f. for wealth for an agrarian society during the communist period. After 32 years of the abolishment of communism the well known Tsallis-Pareto distribution (characteristic for the free market economy) emerged. The LGGR method was able to describe distributions from different periods with kernel functions that are justifiable by the economical policy of the communist and capitalist periods.

We computed and discussed some widely used inequality measures calculated from the estimated wealth data and from the model. We also followed the evolution of these inequality measures for Sâncraiu commune through periods with different socio-economic conditions.

In the final chapter we deal with a somewhat different topic, but we follow a similar research and modelling methodology. We study the Internet as a wired physical network embedded in the physical space. This study is connected to our main topic of wealth and incomes, through the GDP of the cities that influences the growth of the Internet as a network. Using only assumptions based on city size distribution we created an agent-based model, that is capable to generate an Internet-like graph with similar properties. Our model was capable to reproduce a novel scaling between the information spreading time and the geographical distance.

This work fulfilled its original aim as we identified characteristic distributions in socio-economic systems. By approaching these from the direction of complex systems we applied modelling methods to give theoretical explanations to our observations. Our work is only a first attempt in the modelling direction, since we overlook many aspects of the systems in order to obtain a simple description of the observed scaling laws.

For the future we consider important to understand also the subtle connections of our model parameters to the relevant socio-economic parameters. In such manner it would be possible to advertise our work in the social science community, attracting their interest towards the powerful modelling methodologies developed lately by econo-physics.

# List of Publications

## Scientific articles revelant to the thesis

- István Pap, Levente Varga, Mounir Afifi, **István Gere** and Zoltán Néda. *Scaling in the space-time of the Internet.* Scientific Reports, Vol. 9, 9734, (2019), IF = 4.380
- Zoltán Néda, **István Gere**, Tamás S. Biró, Géza Tóth and Noémi Derzsy. *Scaling in income inequalities and its dynamical origin.* Physica A: Statistical Mechanics and its Applications, Vol. 549, 124491, (2020), IF = 3.263
- **István Gere**, Szabolcs Kelemen, Géza Tóth, Tamás S. Biró, and Zoltán Néda. *Wealth distribution in modern societies: Collected data and a master equation approach.* Physica A: Statistical Mechanics and its Applications, Vol. 581, 126194, (2021), IF = 3.263
- **István Gere**, Szabolcs Kelemen, Tamás S. Biró and Zoltán Néda. *Wealth Distribution in Villages. Transition From Socialism to Capitalism in View of Exhaustive Wealth Data and a Master Equation Approach.* Frontiers in Physics, Vol. 10, 827143, (2022), IF = 3.560

## Conference participations

- István Pap, Mounir Afifi, Levente Varga, **István Gere** and Zoltán Néda. (14-17 June 2018) *Nontrivial dynamical scaling in the Internet: experiments and a simple model*. MaCS: 12th Joint Conference on Mathematics and Computer Science, Cluj-Napoca
- István Pap, Mounir Afifi, Levente Varga, **István Gere** and Zoltán Néda. (1-4 May, 2018) *Scaling in the space-time of the Internet*. MECO43: 43th Conference of the Middle European Cooperation in Statistical Physics, Krakow
- **István Gere**, András Kuki and Zoltán Néda. (14-16 September, 2020) *Towards probabilistic forecasting of earthquakes* [Online Conference, Poster presentation]. MECO45: 45th Conference of the Middle European Cooperation in Statistical Physics
- Zoltán Néda, **István Gere** and Szabolcs Kelemen (01-03 July, 2021) *The Growth and Reset model for social inequalities* [Online Conference, Oral presentation]. FENS 2021: 11th Polish Symposium on Physics in Economy and Social Sciences

# Selected bibliography

[1] Schinckus C. *When Physics Became Undisciplined: An Essay on Econophysics (Doctoral thesis).* (Girton College, University of Cambridge, 2018)

[2] Quetelet A. *Lettres à S. A. R. Le Duc Régnant de Saxe Cobourg et Gotha, sur la théorie des probabilités, appliquée aux sciences morales et politique.* (Brussels, Hayez, 1846)

[3] Rice J. A. *Mathematical Statistics and Data Analysis*. (Belmont, CA: Duxbury Press, 2006)

[4] Schmidt K.H. *A new test for random events of an exponential distribution.* European Physical Journal A, Vol. 8, 141–145, (2000)

[5] Néda Z., Varga L., Biró T.S. *Science and Facebook: The same popularity law!* PLoS ONE, Vol. 12, e0179656. (2017)

[6] Gutenberg R., Richter C.F. *Frequency of earthquakes in California.* Bulletin of the Seismological Society of America, Vol. 34, 185-188, (1944)

[7] Rosen K. T., Resnick M. *The size distribution of cities: An examination of the Pareto law and primacy.* Journal of Urban Economics, Vol. 8, 165-186, (1980)

[8] Varga L., Kovács A., Tóth G., Papp I., Néda Z. *Further We Travel the Faster We Go.* PLoS ONE, Vol. 11, e0148913. (2016)

[9] Pareto V. *Cours D'Économie Politique.* (Macmillan, Paris, Vol. 2, 1897)

[10] Kleinert H., Schulte-Frohlinde V. *Critical Properties of* Φ4 *Theories.* (World Scientific Publishing Company, Singapore, 2001.)

[11] Bak P., Tang C., Wiesenfeld K. *Self-organized criticality.* Physical Review A, Vol. 38, 364, (1988)

[12] Newman M. *Power laws, Pareto distributions and Zipf's law.* Contemporary Physics, Vol. 46, 323-351, (2005)

[13] Gere I., Kelemen S., Tóth G., Biró T. S., Néda Z. *Wealth distribution in modern societies: Collected data and a master equation approach* Physica A, Vol. 581, 126194, (2021)

[14] Néda Z., Gere I., Biró T. S., Tóth G., Derzsy N. *Scaling in income inequalities and its dynamical origin.* Physica A, Vol. 549, 124491, (2020)

[15] Biró T.S., Néda Z. *Unidirectional random growth with resetting.* Physica A, Vol. 499, 335-361 (2018)

[16] Pap I., Varga L., Afifi M., Gere I., Néda Z. *Scaling in the space-time of the Internet.* Scientific Reports, 9:9734, (2019)

[17] Abul-Magd A. Y. *Wealth distribution in an ancient Egyptian society.* Physical Review E, Vol. 66, 057104, (2002)

[18] Hegyi G., Néda Z., Santos M.A. *Wealth distribution and Pareto's law in the Hungarian medieval society.* Physica A: Statistical Mechanics and its Applications Vol. 380, 271-277, (2007)

[19] Definition of the word *income*, Cambridge Dictionary, Source: https://dictionary.cambridge.org/dictionary/english/income

[20] Definition of the word: *wealth*, Cambridge Dictionary, Source: https://dictionary.cambridge.org/dictionary/english/wealth

[21] Davies J. B., Shorrocks A. F. *Chapter 11 The distribution of wealth.* Handbook of Income Distribution, Elsevier, Vol. 1, 605-675, (2000)

[22] Derzsy N., Néda Z., Santos M.A. *Income distribution patterns from a complete social security database.* Physica A: Statistical Mechanics and its Applications Vol. 391, 5611–5619. (2012)

[23] Data sources for income, United States of America:
*Current population survey, 2014 annual social and economic supplement/income distribution to $250, 000 or more for males and females: 2013, 2014. PINC-11.-table Income Distribution to $250,000 or More for Males and Females: 2013*, U.S. Census Bureau, (2013), Source: https://www.census.gov/data/tables/time-series/demo/income-poverty/cps-pinc/pinc-11.2013.html

[24] Data sources for income, Finland:
*Finland's PxWeb databases/11bq – Number of income recipients, income estimation, deductions and taxes by income-class*, Statistics Finland's free-of-charge statistical databases, (2017), Source: http://pxnet2.stat.fi/PXWeb/pxweb/en/StatFin/StatFin__tul__tvt/statfin_tvt_pxt_11bq.px/

[25] Data sources for income, Russia:
*Russia in Figures*, Statistical Handbook, M. Rosstat (Ed.), (2017), Source: https://www.gks.ru/free_doc/doc_2017/rusfig/rus17e.pdf

[26] Data sources for income, Australia:
*Taxation statistics 2010–11, Table 14: Individuals tax, Percentile distribution, by taxable income, 2010–11 income year*, Australian Taxation Office Australian Government, (2012), Source: https://www.ato.gov.au/uploadedFiles/Content/CR/Research_and_statistics/In_detail/Downloads/cor00345977_2011IND14.xlsx

[27] Data sources for income, Japan:
*The 141th National Annual Statistics Report FY2015*, National Tax Agency, Japan, Source: https://www.nta.go.jp/publication/statistics/kokuzeicho/h27/h27.pdf
*The 142th National Annual Statistics Report FY2016*, National Tax Agency, Japan, Source: https://www.nta.go.jp/publication/statistics/kokuzeicho/h28/h28.pdf
*The 143th National Annual Statistics Report FY2017*, National Tax Agency, Japan, Source: https://www.nta.go.jp/publication/statistics/kokuzeicho/h29/h29.pdf

[28] Data sources for wealth, United States of America, Russia, France:
World inequality database, data repository, 2021
Source:
https://wid.world/data/ (Accessed 23 February 2021).

[29] Atkinson A.B., Harrison A.J. *Distribution of Wealth in Britain* Cambridge Univ. Press, New York, (1978)

[30] Drăgulescu A., Yakovenko V. *Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States.* Physica A: Statistical Mechanics and its Applications, Vol. 299, 213-221, (2001)

[31] Chakrabarti B.K., Chakraborti A., Chatterjee A. *Econophysics and Sociophysics: Trends and Perspectives.* Wiley-VCH Verlag GmbH & Co. KGaA, (2006)

[32] Dragulescu A., Yakovenko V. *Evidence for the exponential distribution of income in the USA.* European Physical Journal, B 20, 585–589, (2001).

[33] Yakovenko V., and Rosser J. B., *Colloquium: Statistical mechanics of money, wealth, and income.* Reviews of Modern Physics, Vol. 81, 1703-1725, (2009)

[34] *Data repository for the Forbes Rich list, years 2001-2019.* Source : http://datahub.io/gavram/top100-richest-people

[35] Bouchaud J.P., Mézard M. *Wealth condensation in a simple model of economy.* Physica A, Vol. 282, 536–545, (2000)

[36] Biró T.S., Néda Z., Telcs A. *Entropic divergence and entropy related to nonlinear master equations.* Entropy Vol. 21, 993, (2019)

[37] Alvaredo F., Atkinson A., Chancel L., Piketty T., Saez E., Zucman G. *Distributional National Accounts Guidelines: Methods and Concepts Used in the World Inequality Database.* WID working paper, (2020)

[38] Gini C. *On the measure of Concentration with Special Reference to Income and Statistics.* Colorado College Publication, General Series No. 208, 73–79, (1936)

[39] Lorenz M.O., *Methods of measuring the concentration of wealth.* Publications of the American Statistical Association Vol. 9, 209–219, (1905)

[40] Amiri S., Brian R. *Internet penetration and its correlation to gross domestic product: An analysis of the nordic countries.* International Journal of Business, Humanities and Technology Vol. 3, 50–60, (2013)

[41] *Definition of the Internet according to the Encyclopedia Britannica online.*
Source:
https://www.britannica.com/technology/computer/One-interconnected-world

[42] *Computer network communication protocol layers according to the Open System Interconnection (OSI) model.*
Source:
https://www.cloudflare.com/learning/ddos/glossary/open-systems-interconnection-model-osi/

[43] Postel J., *Internet Control Message Protocol*, STD 5, RFC 792, September 1981

[44] *Linux manual page for the "ping" command.*
Source: https://linux.die.net/man/8/ping

[45] Augustin B., Cuvellier X., Orgogozo B., Viger F., Friedman T., et al. *Avoiding traceroute anomalies with Paris traceroute.* IMC 2006 - 6th ACM Internet Measurement Conference, Oct 2006, Rio de Janeiro, Brazil. pp.153-158, (2006)

[46] Internet topology dataset: *The CAIDA IPv4 Routed/24 Topology Dataset – 2017*,
Source:
https://www.caida.org/catalog/datasets/ipv4_routed_24_topology_dataset/

[47] Mahadevan P. et al. *The internet as-level topology: Three data sources and one definitive metric.* ACM SIGCOMM Computer Communication Review (CCR), Vol. 36, Issue 1, 17–26, (2006)

[48] Yook S., Jeong H., Barabasi A. *Modeling the Internet's large-scale topology.* Proceedings of the National Academy of Sciences of the United States of America Vol. 99, 13382–13386, (2002)

[49] Caesar M., Rexford J., *BGP Routing Policies in ISP Networks.* IEEE Network, Vol. 19, (2005)

[50] Zhang-Shen R., McKeown N. *Designing a predictable Internet backbone network.* Proceedings of Third Workshop on Hot Topics in Networks (HotNets-III), Vol. 10, (2004)

[51] Isaac Asimov, *Foundation* Doubleday, (1951)