The size-rank Law and Urban Growth

Frida Alsterlund

Oscar Johander

alfrida@student.chalmers.se
johander@student.chalmers.se

Chalmers University of Technology

Abstract

Since much of economic activity is situated in cities, it is interesting to research the dynamics of urban systems. In this report, the size-rank law, stating that the cities' population and their rank within a nation follows Zipf's law is investigated. We give some examples of real world data, and present an existing model to generate such a distribution. This model has been implemented. The goal of this study is to further expand this model to describe more aspects of the underlying dynamics. More specifically, we incorporate migration of the system's population, based on the growth of the cities. Also, the geographic aspect is studied by letting an external disturbance influence the population's migration. The study shows that this system is tolerant to such expansions, still generating the expected distribution.

1. Introduction

In this report we will study the so called size-rank law in the context of urban growth, giving examples of real world situations of the city size distribution. Then using an existing model, to which we will make some additions and modifications, the generation of such a distribution is examined. The size-rank law states that the sizes and ranks of cities within a nation will follow a power-law distribution. Much of economic activities are centered to take place in cities, and there is a field of research focused on studying the processes through which cities' sizes vary. We will use a network-based sapatial model proposed by Semboloni [1]. Our main interest is to see how we can make this model generate the power-law distribution when it is expanded to include an explicit formulation of migration within the system. We will also investigate how disturbing the systems spatial regions will affect the distribution of the cities.

In the first section, the size-rank law is introduced, and examples of actual size-rank distributions are presented. Then Semboloni's model is presented and implemented. We present our result of this implementation, and the following section list what modifications we have made. The results of running our modified model is presented, and conclusions are drawn from our results. We then list some of the aspects suitable for further investigation.

2. City size distributions: the power law phenomenon

In this section the background of the size-rank law will be introduced. Three real-world examples of the actual size-rank distribution within Sweden, France, and Russia will also be presented.

2.1. Zipf's law

Zipf's law is an empirical law that states that the frequency of a word in a corpus is inversely proportional to its rank. This means that the most frequent word appears approximately twice as often as the second largest, three times as often as the third largest and so on. In other words, the prediction is that a word ranked n occurs $1/n$ times the most frequent word. Figure 1 shows this kind of behavior, and the corresponding logarithmic plot is shown in figure 2.

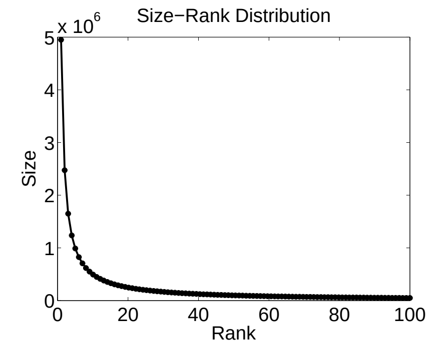

Figure 1: *A theoretic size-rank plot of Zipf's law, showing how the highest ranked dominate.*

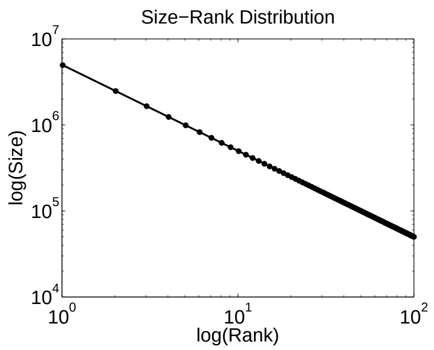

Figure 2: *Thelogarithmic plot of the theoretic size-rank plot in figure 1 showing the power-law behavior.*

The law belong to the group of power law probability distributions and can be described as in equation 1, where one quantity p varies as the power of the other quantity α. The straight line in the logarithmic plot in figure 2 is an indicator for a power law.

$$p \propto r^{-\alpha} \tag{1}$$

Zipf's law apply not only to the frequency of words, but to other cases where ranking occur, such as income rankings and that of city populations. In Oxford's *A dictionary of Geography* the entry on the size-rank law says the following "This 'rule' predicts that, if the settlements in a country are ranked by population size, the population of a settlement ranked n will be $1/n$th of the size of the largest settlement. When settlement size is plotted against rank, on normal graph paper, a concave curve results; plotted on logarithmic scales, a straight line emerges–this is the rank-size pattern". [2]

2.2. Real world examples

In this section a couple of real world examples will be presented to give examples of the actual size-rank distibutions in three countries.

Figure 3 shows the rank-size plot for France and is a good example of how the city population size distribution can be described with a power-law. The data used here is from 2010.[3] Though, the two lowest ranked cities Paris and Marseille deviate some from the predictions.

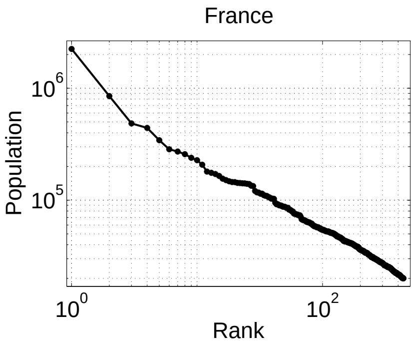

Figure 3: *A size-rank plot over the city population sizes in France. Paris and Marseille, the cities with greatest city population, deviate from the expected power-law distribution but the remaining cities seem to follow it very well.*

Data of the population from 2010 for Swedish cities is displayed in figure 4 [4]. The smaller cities behave as predicted as well as the tree cities with greatest population size, namely Stockholm, Göteborg and Malmö. Though, there is a deviation from the predictions for cities with around 100 000 inhabitants.

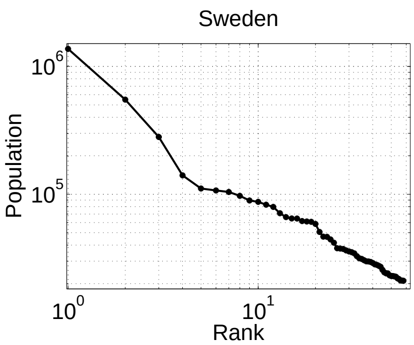

Figure 4: *City population sizes plotted against their rank for cities in Sweden. Note how some of the cities with around 100 000 inhabitants seem to be smaller than expected.*

In Russia, see figure 5, the deviations from a power-law behavior is even larger [5]. The two biggest cities Moscow and Saint Petersburg have a much larger city population than expected and the gap to the next ranked cities is major. This deviation may be a cause of the Sovjet union period, or the size of Russia may contribute to the deviations.

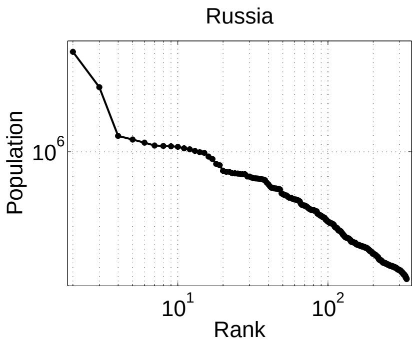

Figure 5: *Size-rank plot for Russia, where the two lowest ranked cities Moscow and Saint Petersburg create a big gap to the next ranked cities because they are so big in comparison.*

A conclusion to draw from the national data over city populations is that the city populations do seem to follow Zipf's law, though there are some deviations too. One problem is how to find appropriate data to investigate. Legal boundaries are not the interesting case when investigating growth, economics rather than geographic boundaries are of more interest here. This may cause some inaccuracy in the data, since it may measure something slightly differently from what we are interested in here.

2.3. Why the size-rank distribution?

The distribution of the city population size in a country is believed to be caused by three contributing factors [6].

- The population size and the population rank are intrinsically negative, which result in a negative slope for the distribution.

- The percental growth is independent, or weakly dependent, of the city population size. Though, the absolute population growth is of course dependent of the population size. This can be compared with Gibrat's law of proportionate growth, that states that the size and the growth rate can be seen as independent of each other and give rise to a log-normal distribution.
- Cities have different growth rates. Natural variations may therefore cause a city to grow a lot for some time, which may be enough for the city to continue to capture demands from other cities. The fact that the growth rates differ between cities results in a nonlinear dynamic.

3. Semboloni's model

In this section, the model used to further investigate urban growth is introduced, we also report our results of our implementation of this model.

3.1. Model description

A model for generating a power law distributed urban system has been proposed by Semboloni [1]. Based on the economic assumptions of a balance between the demand put on a city by nearby smaller cities and its supply proportional to this city's population the dynamics with time generate a power law.

The result can be interpreted as an economic flow from a primary source into smaller and smaller branches of attractions. This is a top-down analogy likened to geographical structures of river networks [7], where this type of hierarchical system arises naturally.

Semboloni instead propose a bottom-up model, and shows analytically that this also generates a power law distributed system. The goods required by a particular city is divided in to two groups: those that can be attained by the city itself, and those which needs to be attained from outside the city. On the other hand, goods supplied by a city is divided into that which is in proportion to the population of the city and that which is not. The latter type can be seen as constituting the top-down flow characterizing the hierarchic structure mentioned above. The largest city supplies goods to smaller nearby cities, which in turn may further distribute even smaller nearby cities, and so on.

In the formulation of the model, the fact that cities are actually scattered on a surface is emphasized and that among the basins of attraction constituted by nearby smaller cities the trading cities are chosen in competition with each other. In order to account for this effect a degree of stochasticity is added both in the way the smaller cities minimize the distance to the nearby cities, and in the way the cities' populations are updated.

3.2. Implementation of model

We have implemented the model as described by Semboloni. The firs step is to randomly scatter N number of cities on a two dimensional surface. The initial populations of the cities are then sampled from a normal distribution g^p centered around a parameter for the average minimal population p_{min} (with standard deviation parameter σ^p). After this comes a procedure in which all cities (except the largest city) find a nearby larger city to get their demand of goods (of the second type mentioned above) satisfied. The cities minimize the distance to the larger cities, using the distance d^*_{ij} between cities i and j calculated as

$$d^*_{ij} = d_{ij}(1 + g^d) \qquad (2)$$

where d_{ij} is the euclidean distance between city i and city j and g^d is sampled from a normal distribution with mean zero and standard deviation σ^d. Given the set of cities, Ω_j that has chosen city j in a particular time-step the population is then updated according to

$$p_j(t+1) = \alpha \sum_{k \in \Omega_j} p_k(t) + p_{min}(1 + g_j^p). \quad (3)$$

Even though this is a linear equation, the inherent stochasticity, and repeated application of this update rule causes the cities to arrange themselves in a distribution that tends to a power law. This is simillar to the effects responsible for the power law mentioned in section 2.3, where slight variations in growth and population increase proportional to the population of a particular city.

The model was executed for $4 \cdot 10^4$ iterations, using the parameters $N = 100$, $\alpha = 0.9$, $\sigma_p = 0.5$, $\sigma_d = 0.1$, and $p_{min} = 1000$. A resulting bubble plot can be seen in figure 6. The sizes of the bubbles are proportional to the cities' populations, and the edges connecting cities indicate the larger city's basin of attraction in the last time step.

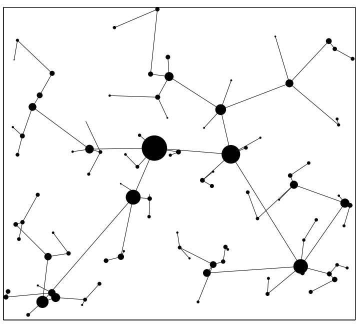

Figure 6: *An example of the cities' configuration after $4 \cdot 10^4$ time steps. The size of the bubbles corresponds to the city's population, and the edges indicate the larger city's basin of attraction. Parameters are: $N = 100$, $\alpha = 0.9$, $\sigma_p = 0.5$, $\sigma_d = 0.1$, and $p_{min} = 1000$.*

The corresponding size-rank distribution can be seen in figure 7. As is evident, the distribution seems to follow a power law for a majority of the largest cities.

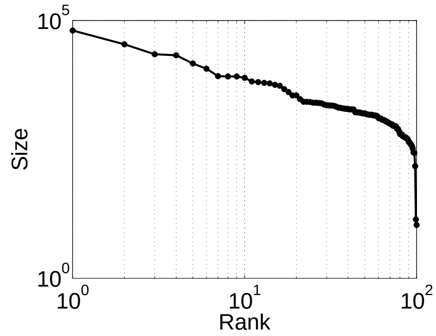

Figure 7: *The size rank plot of the above simulation, after $4 \cdot 10^4$ time steps. Parameters are: $N = 100$, $\alpha = 0.9$, $\sigma_p = 0.5$, $\sigma_d = 0.1$, and $p_{min} = 1000$.*

4. Modified model

In this sections, we review some of the aspects missing in the Semboloni model needed to examine the system at a smaller time scale, and with more variations of the geographic significance.

4.1. Additional aspects

The aspects of the Semboloni model which we wish to use to extend the investigation involve the lack of inertia in the population sizes, making the population dynamics depend on an explicit migration of the population and to futher investigate regional significance other than distance between cities.

The city size of a randomly chosen city is plotted against the time steps in figure 8. It is evident that the size of the city size fluctuates a lot, even between adjacent time steps.The system naturally has memory in terms of the causal effects where some cities might for example accumulate population a couple of subsequent time steps, giving rise to the future developement of the simulation, the lack of inertia where a city is not allowed to change its population size too abruptly makes it difficult to examine various effects on a time-scale shorter than a full simulation run, where the entire system's configuration is considered.

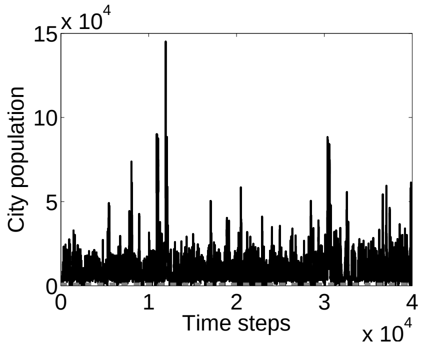

Figure 8: *The size of a randomly choosen city plotted against time steps. The dashed line corresponds to p_{min}. Parameters are: $N = 100$, $\alpha = 0.9$, $\sigma_p = 0.5$, $\sigma_d = 0.1$, and $p_{min} = 1000$.*

Also, in figure 9 the variance of the population up until the time step t, of the randomly chosen city is plotted against the time steps during a full execution of the simulation. The variance indicate that the city population is indeed fluctuating over the entire simulation run, again indicating that the model is more suitible to study the over-all behavior of the system.

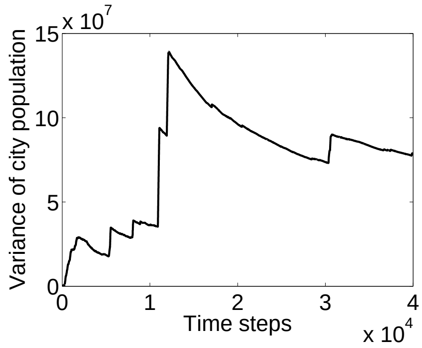

Figure 9: *The variance of the randomly choosen city plotted against time steps, for a full execution of the simulation. Parameters are: $N = 100$, $\alpha = 0.9$, $\sigma_p = 0.5$, $\sigma_d = 0.1$, and $p_{min} = 1000$.*

To account for this fact, and make it possible to study effects on a shorter time scale one possible solution is to instead of interpreting the accumulated demand on a city as directly proportional to its population in the next time step (as in equation 3), to instead make migration within the system explicit.

In addition to adjusting the population dynamics, the regional significance in this model only accounts for the distances between cities in equation 2. The spatial dynamics could be extended to be more dynamic by introduce additional considerations of the cities' positions.

4.2. Growth model

Proceeding from Sembolonis model a new model was created with dynamics depending on the growth rate γ rather than directly being dependent of population size. Each city was instead assigned a fitness value. The fitness value f_j was based on how much city j wanted to grow in an iteration. Equation 3 in Semboloni's model, described in section 3.2, was used to establish this wish for cities to grow or decrease in population size. The growth rate was then found by calculating the difference between wanted size and current population size and normalizing this quantity as in equation 4. The fitness value was distributed over a logarithmic spaced interval between 0.1 and 1, so no city got 0 fitness but cities with more ambition to grow got awarded.

$$\gamma_j = \frac{p_j(t+1) - p_j(t)}{p_j(t)} \tag{4}$$

The ability for people to actually move between cities was also added. A pool of people willing to move in each iteration was created. People were drawn from a normal distribution with mean $q = 0.05$ and $\sigma^q = 0.01$ from every city. This means that approximately 5% of the total population was allowed to move in each time step, but the percental moving from each city differed. A stochastic behavior like that was motivated for because it simulated a reasonable migration. People from the pool were then distributed between cities as a percental proportional to the fitness of the city. The people moving in to a city j, is then calculated as

$$\delta p_j = S \frac{f_j}{\sum_i f_i}, \tag{5}$$

where S denotes the number of people in the pool. With this implementation the size-rank distribution in the Semboloni model was reconceived. Figure 10 shows how the hierarchial structure was preserved. The size-rank plot in figure 11 shows the preserved power law behavior. As in the Semboloni model the growth model ran for 40 000 iterations. Note that the people allowed to move was constrained here because of the pool size. The dynamics are therefore predicted to be slower in the growth model.

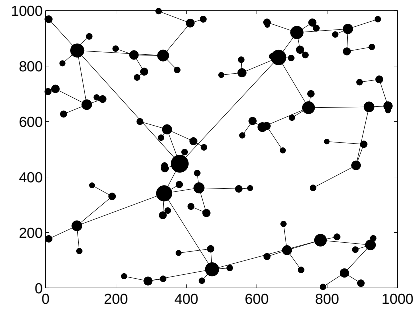

Figure 10: *The hierarchial network of cities retrieved after 40 000 iterations by the model based on growth rate rather than population size directly. Parameters are* N = 100, α = 0.9, σ_p = 0.5, σ_d = 0.1, *and* p_{min} = 1000, $sigma_q$ = 0.01, q = 0.05.

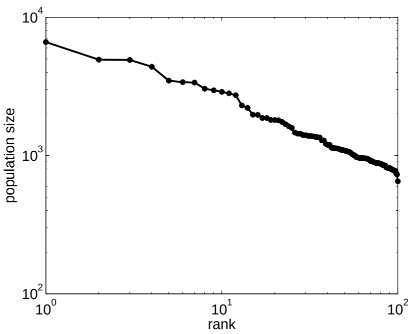

Figure 11: *The size-rank relation in the growth model with added migration. Parameters are* N = 100, α = 0.9, σ_p = 0.5, σ_d = 0.1, *and* p_{min} = 1000, $sigma_q$ = 0.01, q = 0.05.

By basing the dynamics on growth and creating a fitness value it is possible to let the dynamics depend on other aspects then population size directly, which is an improvement compared to Semboloni's model. The system also get some inertia in this model in contrast to the Semboloni model. The model has the advantage that the dynamics of small cities seem more reasonable and decrease fluctuations. Though, for the general behavior of city population growth, Semboloni's model take into consideration enough parameters to simulate the overall behavior of urban city growth.

In figure 12 we see the size of a randomly chosen city plotted against the time steps of a full execution of the simulation. After some initial change in size, the city size seems to maintain a stable value throughout the simulation. Also in figure **??** the variance of this particular city's size is plotted in a similar manner as in figure 9. The variation seem to decrease without interruptions.

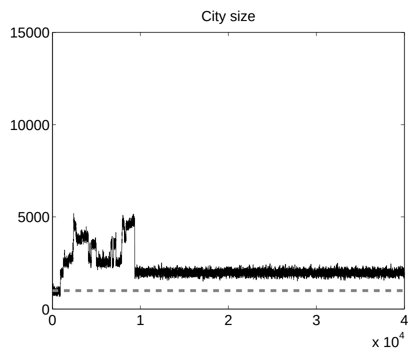

Figure 12: .

4.3. Disturbing the system

To try the robustness of the system and how it react to disruptions, a fitness bonus was added in a region of the map. This disruption could correspond to a sudden change in the economical system. For example findings of natural resources, or a technical innovation, such as an expansion of the railway system.

Figure 13 shows the map of the region. The red cities get a bonus fitness during some iterations to investigate the behavior of the city population sizes on the map. Green cities and black cities in the figure follow the ordinary growth model through all iterations, but are colored differently to easier see the movements in the size-rank plot.

The bonus was calculated by adding a bonus parameter f_{bonus} multiplied by the mean of the fitness values $< f_i >$ and added to the respective towns original fitness, so that

$$f_{j,new} = f_{bonus} < f_i > + f_j \tag{6}$$

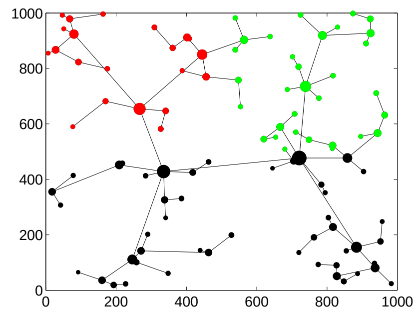

Figure 13: *A map over regions where the red region was favorised in fitness for 20 000 iterations out of 40 000, compared to the green and black region.*

Simulations for 100 cities ran over 10 000 iterations in total, with a fitness bonus $f_{bonus} = 1.25$, beginning with no city given any advantage over the other cities. After stabilization of the system, the regional bonus was put in effect for 75 time steps, meaning that a bonus fitness was added for the red cities in figure 13. The bonus was then removed. The size-rank distributions right before, and after the bonus was in effect can be seen in figure 14, and 15 respectively. Note that the figure of the distribution of the red cities after the bonus was in effect has climbed to lower ranks (larger sizes) while the over all distribution has maintained its shape. The system therefore adapts to the externally changed condition and rearrange its population.

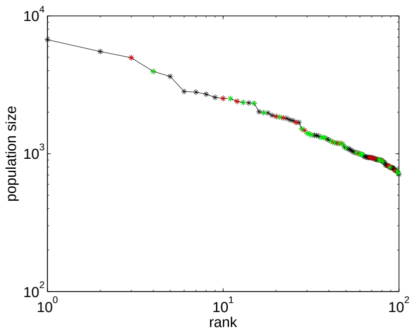

Figure 14: *The size-rank distribution just before the bonus was set in effect.*

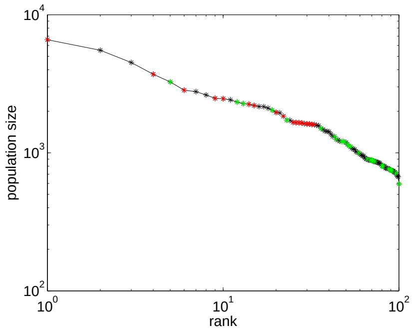

Figure 15: *The size-rank distribution on the final time step of bonus being in effect.*

Figure 16 shows the variation in mean population sizes in the cities that got a bonus and it's neighboring cities. Note how the population among the red cities increses during some time steps to reach a maximum maintained throughout the bonus period, after which the mean population size decreases. This indicates that the external effect had an impact on migration, without disturbing the over all behaviour.

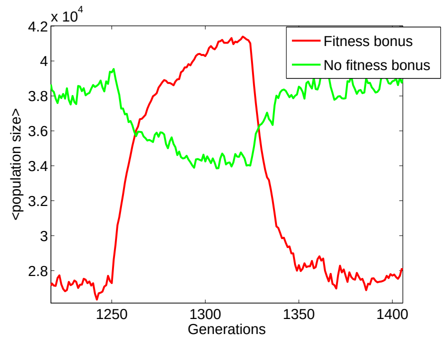

Figure 16: *The variation in mean population sizes in the cities that got a bonus (red) and the neighboring green cities. The bonus was in effect at time step 1250 and disappeared at time step 1325.*

This simulation was subsequently also executed for 40 000 iterations, without any noticeable effect other than those mentioned here observed.

5. Conclusions and discussion

In conclusion, Semboloni's model describes over all generation of a power-law distribution well, but is insufficient to explain effects of population dynamics in great detail. Therefore the model was expanded to take in to consideration a more general description of the growth. This was achieved by explicitly formulating rules for migration of parts of the population of the cities within the system.

The system maintained its power law distribution throughout the simulations, and exhibited stable population dynamics. In addition, the system was regionally disturbed while the migration of the population was monitored, with the result that the system dynamically adapted to these disturbances while maintaining its properties.

The way the growth model was implemented, using fitness values logarithmically spaced over the cities needs to be further investigated, to assess its implications on the model. In a similar manner, the way the bonus awarded to a particular region of the surface as in equation 6. This could probably be implemented in a more realistic way.

5.1. Further investigations

Several steps can be taken to futher expand the scope of effects that could be investigated in this model. For example, in our model the total population is constant, depending on the initial randomization of the cities' populations. This is also true in the Semboloni model, and in both cases the total number of inhabitants is strongly corelated to the parameter p_{min}. Another possibility would be to let the total population grow in accordance to some percental annual growth rate, based on real-world data, accounting for child birth and immigration. This could be technically solved by at each time step increasing the pool by a factor corresponding to annual percental growth.

In addition, the dynamics of new cities arising and letting cities with declining population go extinct could be incorporated in to the model. For example, in the investigations of regional growth, the flux of people into a particular region could give rise to a new town estabilishing given some threshold. Similary, if the people moving from a particular city (with a small population) in some number of adjacent time steps, rises over a particular threshold the city could go extinct.

References

[1] Semboloni, F. (2008) Hierarchy, cities size distribution and Zipf's law. *European Physical Journal B*, 63, nr. 3, pp. 295-301

[2] Mayhew, Susan (ed.) Rank-size rule (2009) *A Dictionary of Geography (4 ed.)* : Oxford University Press

[3] INSEE: Institute National de la Statistique et des Études (2010) www.insee.fr

[4] SCB: Statistics Sweden (2010) *Localities 2010* `www.scb.se` (2013-12-12)

[5] Federal State Statistics Service (2011) *2010 All-Russia Population Census* `www.gks.ru` (2013-12-12)

[6] Statictical Concultants Ltd (2011) *The Rank-Size Rule of City Populations* `www.statisticalconsultants.co.nz` (2013-12-10)

[7] Woldenberg, J.M., Berry B.J.L. (1967) Rivers and central places: analogous systems? Journal of regional science, vol. 7 nr. 2, pp. 129-139 1967