
Weighted networks at the Polish market

A. M. Chmiel, J. Sienkiewicz, K. Suchecki, and J. A. Hołyst

Faculty of Physics and Center of Excellence for Complex Systems Research
Warsaw University of Technology, Koszykowa 75, PL 00-662 Warsaw, Poland
jholyst@if.pw.edu.pl

1 Introduction

During the last few years various models of networks [1, 2] have become a powerful tool for analysis of complex systems in such distant fields as Internet [3], biology [4], social groups [5], ecology [6] and public transport [7]. Modeling behavior of economical agents is a challenging issue that has also been studied from a network point of view. The examples of such studies are models of financial networks [8], supply chains [9, 10], production networks [11], investment networks [12] or collective bank bankruptcies [13, 14]. Relations between different companies have been already analyzed using several methods: as networks of shareholders [15], networks of correlations between stock prices [16] or networks of board directors [17]. In several cases scaling laws for network characteristics have been observed.

In the present study we consider relations between companies in Poland taking into account common branches they belong to. It is clear that companies belonging to the same branch compete for similar customers, so the market induces connections between them. On the other hand two branches can be related by companies acting in both of them. To remove weak, accidental links we shall use a concept of threshold filtering for weighted networks where a link weight corresponds to a number of existing connections (common companies or branches) between a pair of nodes.

2 Bipartite graph of companies and branches

We have used the commercial database "Baza Kompas Polskie Firmy B2B" [18] from September 2005. It contains information concerning over 50 000 large and medium size Polish companies belonging to one or more of 2150 different branches. A bipartite graph of companies and branches has been constructed as at Fig. 1.

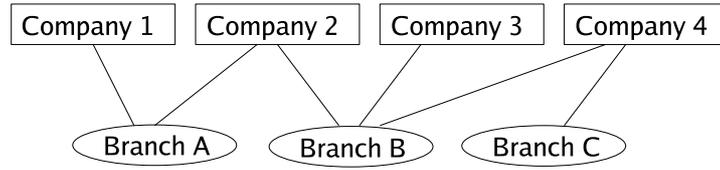


Fig. 1. Bipartite graph of companies and trades.

In the bipartite graph we have two kinds of objects: branches $A = 1, 2, 3, \dots, N_b$ and companies $i = 1, 2, 3, \dots, N_f$, where N_b – total number of branches and N_f – total number of companies. Let us define a *branch capacity* $|Z(A)|$ as the cardinality of set of companies belonging to the branch A . At Fig. 1 the branch A has the capacity $|Z(A)| = 2$ while $|Z(B)| = 3$ and $|Z(C)| = 1$. The largest capacity of a branch in our database was 2486 (construction executives), the second largest was 2334 (building materials).

Let $B(i)$ be a set of branches a given company i belongs to. We define a *company diversity* as $|B(i)|$. An average company diversity μ is given as

$$\mu = \frac{1}{N_f} \sum_{i=1}^{i=N_f} |B(i)|. \quad (1)$$

For our data set we have $\mu = 5.99$.

Similarly an average branch capacity ν is given as

$$\nu = \frac{1}{N_b} \sum_{A=1}^{A=N_b} |Z(A)|, \quad (2)$$

and we have $\nu = 134$.

It is obvious that the following relation is fulfilled for our bipartite graph:

$$\mu N_f = \nu N_b. \quad (3)$$

3 Companies and branches networks

The bipartite graph from Fig. 1 has been transformed to create a *companies network*, where nodes are companies and a link means that two connected companies belong to at least one common branch. If we used the example from Fig. 1 we would obtain a companies network presented at Fig. 2.

We have excluded from our dataset all items that correspond to communities (local administration) and as a result we got $N_f = 48158$ companies, belonging to a single connected cluster. Similarly a *branch network* has been

constructed where nodes are branches and an edge represents connection if at least one company belongs to both branches. In our database we have $N_b = 2150$ different branches.

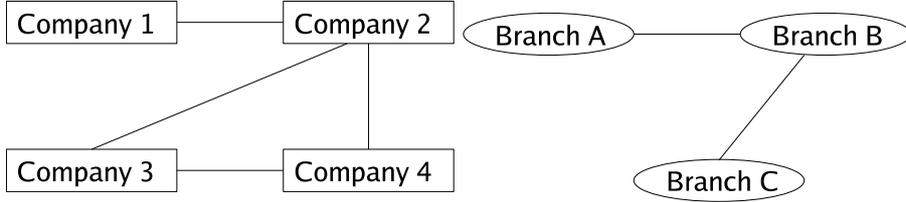


Fig. 2. Companies network on the left, branches network on the right.

4 Weight, weight distribution and networks with cutoffs

We have considered link-weighted networks. In the branches network the link weight means a number of companies that are active in the same pair of branches and it is formally a cardinality of a common part of sets $Z(A)$ and $Z(B)$, where $Z(A)$ is a set of companies belonging to the branch A and $Z(B)$ is a set of companies belonging to the branch B :

$$w_{AB} = |Z(A) \cap Z(B)|. \quad (4)$$

Let us define a function f_k^A which is equal to one if a company k belongs to the branch A , otherwise it is zero:

$$f_k^A = \begin{cases} 1, & k \in A \\ 0, & k \notin A \end{cases}. \quad (5)$$

Using the function f_k^A the weight can be written as:

$$w_{AB} = \sum_{k=1}^{N_F} f_k^A f_k^B. \quad (6)$$

The weight distribution $p(w)$, meaning the probability p to find a link with a given weight w , is presented at Fig. 3. The distribution is well approximated by a power function

$$p(w) \sim w^{-\gamma}, \quad (7)$$

where the exponent $\gamma = 2.46 \pm 0.07$. One can notice the existence of edges with large weights. The maximum weight value is $w_{max} = 764$, and the average weight

$$\langle w \rangle = \sum_{w_{min}}^{w_{max}} wp(w), \quad (8)$$

equals $\langle w \rangle = 4.67$.

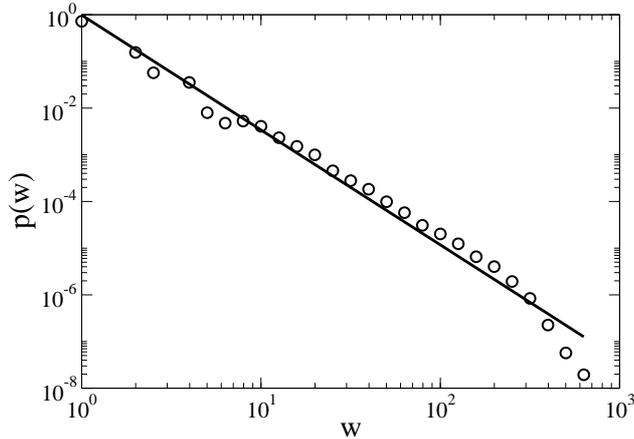


Fig. 3. Weight distribution in branches network.

Using cutoffs for link weights we have constructed networks with different levels of filtering. In such networks nodes are connected only when their edge weight is no less than an assumed cutoff parameter w_o . In Table 1 we present the main parameters of branches networks with changing w_o . The clustering coefficient has been calculated by the definition:

$$C = \frac{1}{N} \sum_{i=1}^N \frac{2E_i}{k_i(k_i - 1)} \quad (9)$$

where E_i is the number of existing connections between neighbours of node i and k_i is the degree of vertex i . The values of clustering coefficient is very high in comparison to random networks, and it remains almost at a constant level for the presented values of cutoff parameter.

A weight in the companies network is defined in a similar way as in the branches networks, it is the number of common branches for two companies — formally it is equal to the cardinality of a common part of sets $B(i)$ and $B(j)$, where $B(i)$ is a set of branches the company i belongs to, $B(j)$ is a set of branches the company j belongs to:

$$w_{ij} = |B(i) \cap B(j)|. \quad (10)$$

Table 1. Data for branches networks: w_o is the value of selected weight cutoff, N is the number of vertices with nonzero degrees, E is the number of links, k_{max} is the maximum node degree, $\langle k \rangle$ is the average node degree, C is the clustering coefficient.

w_o	N	E	k_{max}	$\langle k \rangle$	C
1	2150	389542	1716	362	0.530
2	2109	212055	1381	201	0.565
3	2053	136036	1127	132	0.568
4	2007	100917	952	100	0.575
5	1948	80358	802	82	0.589
6	1904	66353	655	69	0.592
7	1858	56565	569	60	0.596
8	1819	49193	519	54	0.597
9	1786	43469	477	48	0.599
10	1748	38924	450	44	0.600
12	1666	32167	394	38	0.615
14	1611	26088	325	32	0.605
16	1545	21762	288	28	0.606
18	1490	18451	259	24	0.603
20	1424	15872	226	22	0.604
30	1188	8989	162	15	0.585
40	996	6036	131	12	0.587
50	857	4379	111	10	0.572
60	752	3303	85	8	0.551
70	666	2638	65	7	0.524
80	575	2143	55	7	0.532
90	512	1808	49	7	0.538
100	464	1543	41	6	0.546
150	306	750	26	4	0.493

Using the function f_k^A the weight can be written as

$$w_{ij} = \sum_{A=1}^{N_b} f_i^A f_j^A. \quad (11)$$

The maximum value of observed weights $w_{max} = 207$ in this network is smaller than in the branches network while the average value equals $\langle w \rangle = 1.48$. The weight distribution in this case does not follow a power law and in a limited range it shows an exponential behavior.

Similarly to the branches networks we have introduced cutoffs in companies network. The dependence of selected network parameters on cutoff threshold is shown in Table 2. The behaviour of clustering coefficient resembles the one observed in the branches networks.

At Fig. 4 we present average degrees of nodes and maximum degrees as functions of the cutoff parameter w_o . We have observed a power law scaling

$$\langle k \rangle \sim w_o^{-\beta}, \quad (12)$$

Table 2. Data for companies networks: w_o is the selected cutoff, N is the number of nodes with nonzero degrees, E is the number of links, k_{max} is the maximum node degree, $\langle k \rangle$ is the average node degree, C is the clustering coefficient.

w_o	N	E	k_{max}	$\langle k \rangle$	C
1	48158	39073685	16448	1622	0.652
2	39077	9932790	8366	508	0.689
3	31150	3928954	4842	252	0.714
4	24212	1895373	3103	156	0.717
5	18566	1024448	2059	110	0.713
6	14116	622662	1412	88	0.710
7	10796	404844	1012	74	0.700
8	8347	266013	724	63	0.701
9	6527	180696	566	55	0.699
10	5197	124079	443	47	0.699
11	4268	94531	382	44	0.704
12	3400	68648	345	40	0.693
13	2866	54258	305	37	0.691
14	2277	36461	277	32	0.663
15	1903	28844	249	30	0.673
16	1627	23063	231	28	0.678
17	1397	18352	212	26	0.667
18	1196	14480	191	24	0.680
19	1003	11230	171	22	0.680
20	883	8907	159	20	0.676

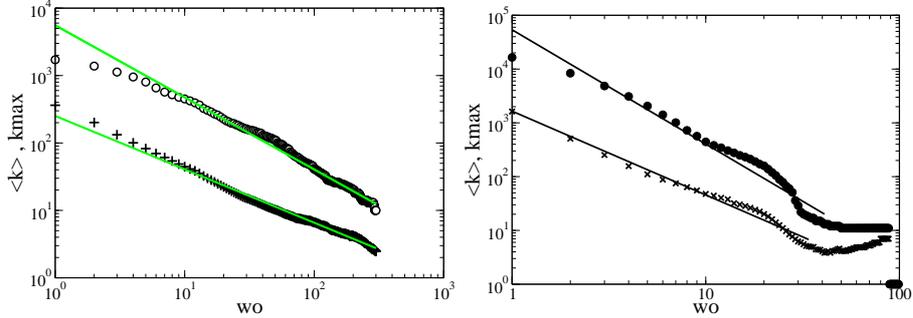


Fig. 4. Dependence of $\langle k \rangle$ and k_{max} on cutoff parameter w_o for branches networks (left) and companies networks (right).

$$k_{max} \sim w_o^{-\alpha}, \quad (13)$$

where for branches networks $\alpha_b = 1.069 \pm 0.008$ and $\beta_b = 0.792 \pm 0.005$ while for companies networks $\alpha_f = 2.13 \pm 0.07$ and $\beta_f = 1.55 \pm 0.04$.

5 Degree distribution

We have analyzed the degree distribution for networks with different cutoff parameters. At Fig. 5 we present the degree distributions for companies networks for different values of w_o . The distributions change qualitatively with increasing w_o from a nonmonotonic function with an exponential tail (for $w_o = 1$) to a power law with exponent γ (for $w_o > 6$).

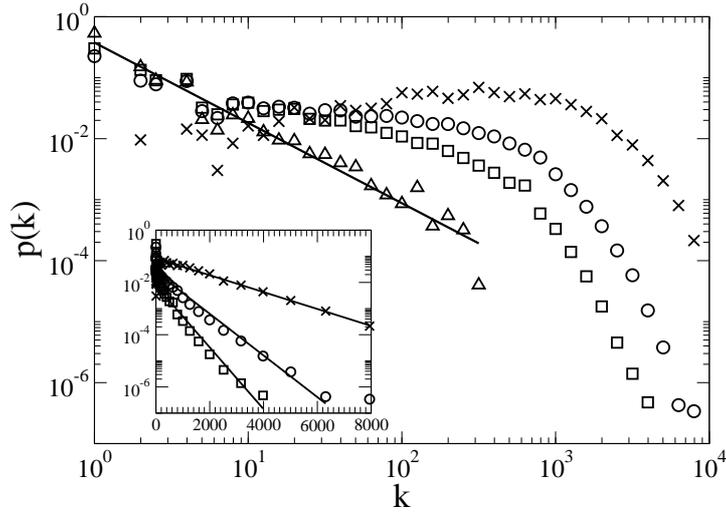


Fig. 5. Degree distributions for companies networks for different values of w_o . X-marks are for $w_o = 1$, circles are for $w_o = 2$, squares are for $w_o = 3$ and triangles are for $w_o = 12$.

At the Fig. 6 we present a degree distribution for branches networks. For $w_o = 1$ we observe a high diversity of node degrees — vertices with large values of k occur almost as frequent as vertices with a small k . For a properly chosen cutoff values the degree distributions are described by power laws. For $w_o = 4$ we see two regions of scaling with different exponents γ_1 and γ_2 while a transition point between both scaling regimes appears at $k \approx 100$. Branches belonging to the first regime of scaling are more specific, for example "production of protective clothing", "poultry farmer", "gemstone" and branches on the right are more general like "import and export general", "network of supermarkets". We suppose that the mechanism of this behaviour is similar to linguistic networks [19]. Appearance of a new general branch creates connections between existing specific branches, what causes double scaling in the end.

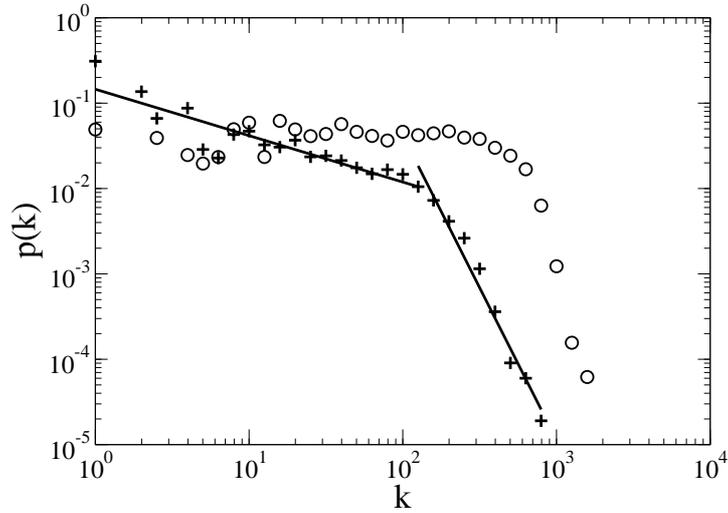


Fig. 6. Degree distribution in branches network for different values of w_o . Circles are for $w_o = 1$, crosses are for $w_o = 4$.

It is important to stress that in both networks (companies and branches) the scaling behavior for degree distribution occurs only if we use cutoffs for links weights (compare Fig. 5 and Fig. 6). It follows that such cutoffs act as filters for the noise present in the complex network topology.

6 Entropy of network topology

In order to examine how much information we filter out, we have chosen to investigate entropy S , using a standard formula for Gibbs entropy (14):

$$S = - \sum_k p(k) \ln p(k). \quad (14)$$

The entropy of degree distribution in both networks decays logarithmically as a function of the cutoff value w_o (see Fig. 7):

$$S = -a \ln(w_o) + b. \quad (15)$$

For branches networks fitting parameters are $a = 0.834 \pm 0.004$ and $b = 6.51 \pm 0.02$.

The entropy in companies networks behaves similarly with $a = 1.79 \pm 0.05$ and $b = 8.49 \pm 0.15$. We have observed that decrease is faster with w_o than in the branches networks. This is due to different distributions of weight in both

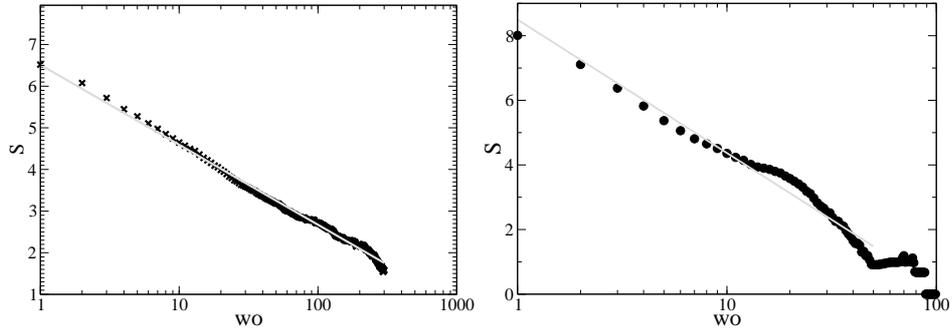


Fig. 7. Entropy dependence on cutoff parameter for branches networks on the left and for companies networks on the right.

networks and a smaller range of weights in the companies networks. Since they are filtered differently with w_o we have observed dependence of S on the parameter $\langle k \rangle$. This dependence is presented at the Fig. 8 along with results of formula (16) with real values of k_{max} and γ . The formula (16) is calculated for general case of power law distributions with $k_{min} = 1$.

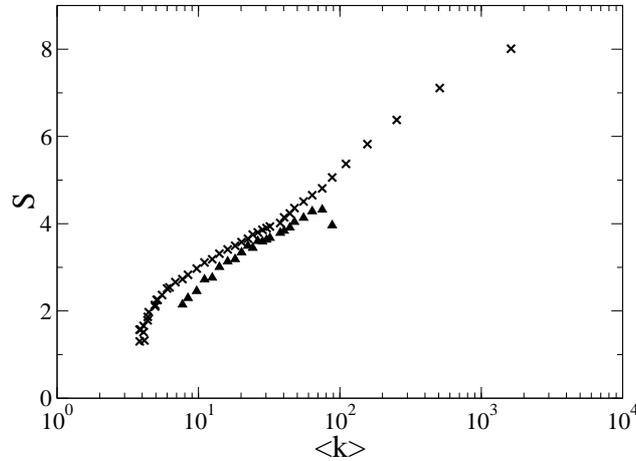


Fig. 8. Entropy dependence on the average node degree $\langle k \rangle$. X-marks represent measured entropy in companies networks and triangles represent the results of analytic formula (16) (only for networks with definite power-law distribution).

$$S(\gamma, k_{max}) = -\ln \frac{1-\gamma}{k_{max}^{(1-\gamma)} - 1} + \frac{\gamma k_{max}^{1-\gamma} \ln k_{max}}{k_{max}^{(1-\gamma)} - 1} - \frac{\gamma}{(1-\gamma)} \quad (16)$$

In our system, parameters γ and k_{max} depend on each other. Both are defined by the cutoff parameter w_o . The increase of w_o results in the decrease of k_{max} and increase of γ . Let us take the simplest possible situation — a uniform distribution from $k_{min} = 1$ to k_{max} . For this case we calculated the entropy analytically using the definition (14):

$$S(k_{max}) = -\int_1^{k_{max}} \frac{1}{k_{max} - 1} \ln\left(\frac{1}{k_{max} - 1}\right) dk = \ln(k_{max} - 1), \quad (17)$$

what is equivalent to :

$$S = \ln(\langle k \rangle - 1) + \ln 2. \quad (18)$$

The value of $\langle k \rangle$ is strictly connected to the width of the distribution. Figure 9

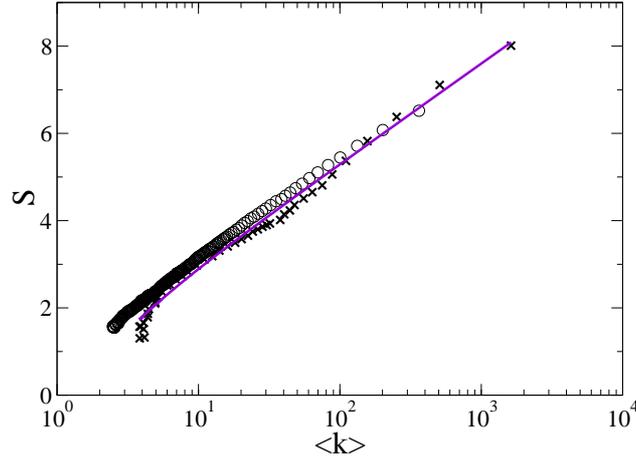


Fig. 9. Dependence of entropy on the average nodes degree. Circles represent branches networks and X-marks represent companies networks. Line corresponds to (18).

shows that such a simplistic approach gives a very good approximation of real entropy value. We can conclude that the width of the distribution is the main source of entropy changes in our systems. The presence of two parameters (γ, k_{max}) is irrelevant as both are derived from the same parameter w_o . We decided to use the single parameter $\langle k \rangle$ since it seems to be the most significant and universal for different networks.

7 Conclusions

In this study, we have collected and analyzed data on companies in Poland. 48158 medium/large firms and 2150 branches form a bipartite graph that allows to construct weighted networks of companies and branches.

Link weights in both networks are very heterogenous and a corresponding link weight distribution in the branches network follows a power law. Removing links with weights smaller than a cutoff (threshold) w_o acts as a kind of filtering for network topology. This results in a recovery of a hidden scaling relations present in the network. The degree distribution for companies networks changes with increasing w_o from a nonmonotonic function with an exponential tail (for $w_o = 1$) to a power law (for $w_o > 6$). For a filtered ($w_o > 4$) branches network we see two regions of scaling with different exponents. Entropies of degree distributions for both networks decay logarithmically as a function of cutoff parameter and are proportional to the logarithm of the mean node degree. We have found the distribution width to be a crucial factor for entropy value.

8 Acknowledgements

We acknowledge a support from the EU Grant *Measuring and Modeling Complex Networks Across Domains* — MMCOMNET (Grant No. FP6-2003-NEST-Path-012999) and from Polish Ministry of Science and Higher Education (Grant No. 13/6.PR UE/2005/7), and from a special grant of Warsaw University of Technology.

References

1. Albert R, Barabási A-L (2002) Statistical mechanics of complex networks, *Reviews of Modern Physics* 74:47-97
2. Newman M E J (2003) The structure and function of complex networks, *SIAM Review* 45:167-256
3. Pastor-Satorras P, Vespignani A (2004) *Evolution and Structure of the Internet: A Statistical Physics Approach*, Cambridge University Press, Cambridge
4. Ravasz E, Somera A L, Mongru D A, Oltvai Z N, Barabasi A-L (2002) Hierarchical organization of modularity in metabolic networks, *Science* 297:1551-1555
5. Newman M E J, Park J (2003) Why social networks are different from other types of networks, *Physical Review E* 68:036122
6. Garlaschelli D, Caldarelli G, Pietronero L (2003) Universal scaling relations in food webs, *Nature* 423:165-168
7. Sienkiewicz J, Hołyst J A (2005) Statistical analysis of 22 public transport networks in Poland, *Physical Review E*, 72:046127
8. Caldarelli G, Battiston S, Garlaschelli D, Catanzaro M (2004) Emergence of Complexity in Financial Networks. In: Ben-Naim E, Frauenfelder H, Toroczkai Z (eds) *Lecture Notes in Physics* 650:399 - 423, Springer-Verlag

9. Helbing D, Lämmer S, Seidel T (2004) Physics, stability and dynamics of supply networks, *Physical Review E* 70:066116
10. Helbing D, Lämmer S, Witt U, Brenner T (2004) Network-induced oscillatory behavior in material flow networks and irregular business cycles, *Physical Review E*, 70:056118
11. Weisbuch G, Battiston S (2005) Production networks and failure avalanches, e-print physics/0507101
12. Battiston S, Rodrigues J F, Zeytinoglu H (2005) The Network of Inter-Regional Direct Investment Stocks across Europe, e-print physics/0508206
13. Aleksiejuk A, Hołyst J A (2001) A simple model of bank bankruptcies, *Physica A*, 299:198-204
14. Aleksiejuk A, Hołyst J A, Kossinets G (2002) Self-organized criticality in a model of collective bank bankruptcies, *International Journal of Modern Physics C*, 13:333-341
15. Garlaschelli G, Battiston S (2005) The scale-free topology of market investments, *Physica A*, 350:491-499
16. Onnela J-P, Chakraborti A, Kaski K, Kertész J, Kanto A (2003) Dynamics of market correlations: Taxonomy and portfolio analysis, *Physical Review E*, 68:056110
17. Battiston S, Catanzaro M (2004) Statistical properties of corporate board and director networks, *European Physical Journal B* 38:345-352
18. See web page of company <http://www.kompass.com>
19. Dorogovtsev S N, Mendes J F F (2001) Language as an evolving word web, *Proceedings Royal Society of London Series B, Biological Sciences* 268 (1485): 2603-2606