Network Structure of Japanese Firms. Hierarchy and Degree Correlation: Analysis from 800,000 Firms

Tomohiko Konno

Graduate School of Economics, University of Tokyo
Research Fellow of the Japan Society for the Promotion of Science

Abstract

We analyzed fundamental characteristics of transaction network by using Japanese 800,000 firms' data. We found hierarchical structure and negative degree correlation in firms' transaction network. The network consists of 800,000 Japanese firms. We also summarize other features as to network and discuss why studying network structure is important. We also found scale free distribution in undirected network.

JEL Classification: L16
Keywords: Network of firms; Scale free network; Complex network; Hierarchy; Degree correlation
Correspondence: Bunkyo-ku, Hongo 7-3-1 Tokyo, Japan
E-mail: tomo.konno@gmail.com

The author acknowledges Hiroshi Yoshikawa and Research Institute of Economy, Trade and Industry(RIETI). This research was partially supported by JSPS.

1 Introduction

We studied Japanese firms' transaction network. We revealed that the network has hierarchy and degree correlation. We discovered hierarchy by analyzing clustering coefficient. We also discovered scale free degree distribution in undirected network the meaning of which will be explained later.

Garlaschelli, Battiston, Castri, Servedio and Caldarelli (2005); Souma, Fujiwara and Aoyama (2006) studied shareholders network in US, Japan and Italy. They found scale free characteristics. However their data consist of only small number of firms. On the other hand, our data contains 800,000 firms. Saito, Watanabe and Iwamura (2007) used the same data of us and revealed that directed network has scale free distribution. We believed that there still remained some important characteristics as to network, then we discovered other significant results. Our motivation comes from the raise of complex network theory and the belief that network is also important to Economics not only for other fields of science. Rather, we believe that view of complex networks can have much influence on Economics. The reviews as to complex network are as follows S.N.Dorogovtesev and J.F.F.Mendes (2003); Vega-Redondo (2007).

The purpose of the present paper is, mainly, to give information of firm's transaction network for the future study of models on network, for example, competing firms on network or so. Unless we know about real network structure, we do not have any idea what kind of network we can assume. Speaking of models on network, it has been already known that in evolutionary games literature network structure affects results of model. For example, Abramson and Kuperman (2001) discussed that Prisoner's Dilemma evolutionary games, in short PD game, differ across network structures ranging from regular lattices to random networks. Santos and Pacheco (2005) discussed cooperative behavior in PD game is enhanced in scale free network. Ohtsuki, Hauert, Lieberman and Nowak (2006) discussed the rule of enhancing cooperation for evolutionary games on network. Hence, in Economics it is likely that models on network are dependent on underlying network structures. This paper will lead to such kinds of studies in the near future.

1.1 Models on Network

In complex network theory, as is mentioned above, they discovered that there are significant relations between behavior of agents on the network and underlying network structure mainly in evolutionary games (Ohtsuki et al. (2006); Santos and Pacheco (2005); Abramson and Kuperman (2001)). This kind of relation between games and network is one of the motivations which drives us to study real network structure. As to hierarchical structure, Jeromos and Gyorgy (2005) discussed that PD games on hierarchical regular lattice, the highest frequency of cooperation occurs in the middle layers, if there are enough layers. They discussed optimum number of layers for the community. They also discussed cooperation of PD game is diminished on hierarchical scale free network. Aoki and Yoshikawa (2006) discussed the importance of hierarchical structure in Finance and Economics.

Speaking of degree correlation, co-star and co-author relations have positive degree correlation, on the other hand, gene network, protein network, nerve circuit and food chain, that appear in Biology, have negative correlation. Artificial network like power grid and Internet have negative but weak correlation. Degree correlation differs across different networks. Positive correlation tends to lower percolation transition point Newman (2002); Callaway, Hopcroft, Kleinberg, Newman and Strogatz (2001). In particular, it is well known that degree correlation affects synchronization of oscillators on the network Motter, Zhou and Kurths (2005); Di Bernardo, Garofalo and Sorrentino (2007); Sorrentino, di Bernardo, Cuellar and Boccaletti (2006); di Bernardo, Garofalo and Sorrentino (2005). In this sense, degree correlation is an important feature of the network.

1.2 Outline of this paper

The outline of this paper is as follows. Section 2 consists of three subsections. The first subsection shows the approach to identifying hierarchical structure and degree correlation. It consists of four blocks. First, we explain the figure which we will use many times. Second, we briefly introduce random network and scale free network. Third, we describe what is hierarchical structure by comparing other kinds of network which does not have hierarchy and approach to identifying hierarchical structure. Finally, we introduce degree correlation. Then the second subsection deals with the following three blocks. First, we describe the data. Second, we explain the way of calculating degree distribution. And the last, we introduce clustering coefficient. In the last subsection, we show the results and discussions in three blocks. First, we demonstrate degree distribution of the Japanese firms' undirected network. Second, we show the network has hierarchical structure by analyzing clustering coefficient. And the last, we show the network has negative degree correlation. Section 3 is the conclusion.

2 Analysis

2.1 The approach to identifying hierarchical structure and degree correlation

First of all, we must explain Fig.1[1] briefly. There are three kinds of network: Random network, Scale free network and hierarchical network. We will explain in this order. In the figure, k stands for degree. The definition of degree is the number of links the vertex has. The first row(a) is for example of each network, the second row(b) is for degree distribution $P(k)$, and the third row(c) is for clustering coefficient $C(k)$. The meaning of the later two terms will be explained.

2.1.1 Scale Free Network and Random Network

First, let us introduce scale free network. Scale free network is a network whose degree distribution follows,

$$P(k) \sim k^{-\gamma} \tag{1}$$

Degree means how many links a vertex has and k is used to stand for

[1] Reprinted by permission from Macmillan Publishers Ltd: [Nature Reviews Genetics] Barabasi and Oltvai (2004), copyright(2004)

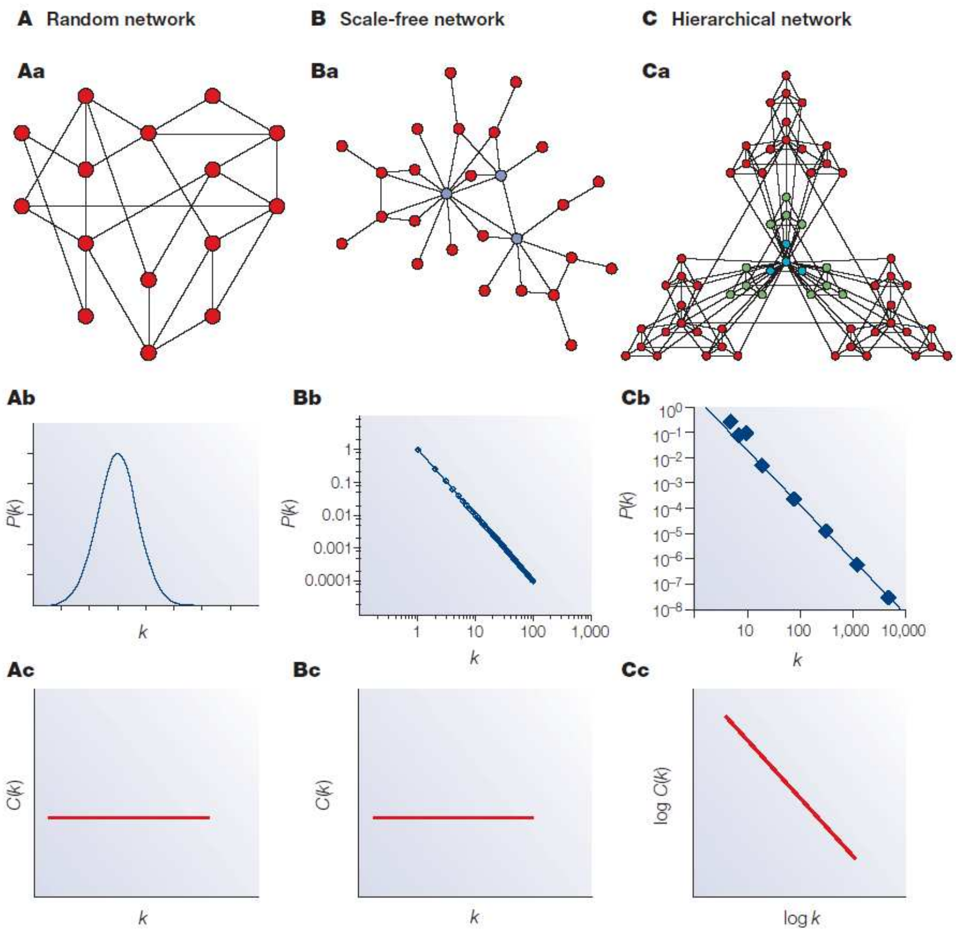

Figure 1: Network Structures

degree. Scale free network is quite different from "Random Graph", which was initiated by P.Erdos and A.Renyi (1959). Random graph is constructed in the following way. Choose two vertices and link them in probability p or, in other words, do not link them in $1-p$. Complete this procedure for all pairs of vertices. If there are n vertices on the whole network, the degree distribution is Binomial,

$$\begin{aligned} p(k) &= {}_{n-1}C_k\, p^k (1-p)^{n-1-k} \\ &\sim {}_nC_k\, p^k (1-p)^{n-k} \quad (\because n-1 \sim n) \end{aligned} \tag{2}$$

In the limit, $n \to \infty, p \to 0$ with keeping $np = \lambda$, eq.(2) becomes Poisson distribution as

$$P(k) = \frac{e^{-\lambda}\lambda^k}{k!} \tag{3}$$

λ is mean degree of the network. Aa and Ab in Fig.1 illustrate random graph and its degree distribution.

On the other hand, Ba and Bb in Fig.1 show scale free network and its degree distribution in log-log plot. On random graph, there are no vertices which have very large degree. On scale free network, in contrast, there are small number of vertices which have very large degree. We call them "Hub". Roughly speaking, the existence of "Hub" is the difference between two networks.

2.1.2 The approach to identifying hierarchical structure

The third network(C) is hierarchical network, which also has scale free degree distribution such as $P(k) \sim k^{-\gamma}$. The difference between scale free network(B) and hierarchical network depicted in the third row(Bc,Cc) which illustrate clustering coefficients for these networks. Clustering coefficient, we will explain later, of scale free network is constant $C(k) = \text{Const}$, however on the other hand that of hierarchical network is dependent on k as $C(k) \sim k^{-1}$. To compare this hierarchical network with other two types of network, it becomes clear what is like hierarchical network structure. In many real networks, clustering coefficient and degree have the above special relation, $C(k) \sim k^{-1}$. To raise some examples, Co-Actor network, Language network, World Wide Web and Internet at the autonomous system level. Ravasz and Barabasi (2003) showed that this relation, $C(k) \sim k^{-1}$, implies hierarchical structure. The authors called hierarchical structure like (Cc) in Fig.1. Barabasi and Oltvai (2004) as well explains this relation and hierarchical structure.

As is explained, in graph Cc $\log c(k)$ and $\log k$ are linearly proportional and the proportionality coefficient is -1, while clustering coefficient $C(k)$ is constant in other two networks: Random network(A) and Scale free network(B). Therefore we studied clustering coefficient to detect the relation $C(k) \sim k^{-1}$ which implies hierarchical structure like Ca[2] in Fig.1.

2.1.3 Degree Correlation

We also study degree correlation of the network. Degree correlation is defined by the following equation

$$k_{nn}(k) \equiv \sum_{k'} k' \Pr(k' \mid k) \tag{4}$$

$\Pr(k' \mid k)$ is the conditional probability that the vertex with degree k is adjacent to the vertex degree of which is k'. In a nutshell, $k_{nn}(k)$ means that the vertex with degree k is adjacent to the vertex with degree $k_{nn}(k)$. This is also an important network characteristic.

2.2 The data, degree distribution and clustering coefficient

2.2.1 The data

The data is supplied by Tokyo Shoko Research, Ltd(TSR) via RIETI. The data consists of 800,000 firms' financial data and network relationships: buy, sell and shareholder relationship. The data reports 4,000,000 relations and includes information of firms such as gross sales, region, year built, the num-

[2]similar figure appears in Ravasz and Barabasi (2003)

ber of employee, the number of office, the number of factories, industrial classification and so on. We do not make use of any firm which does not report gross sales, subsequently the number of firms become 800,000. We used relation of buy and sell. This data set was made by asking firms to raise their business partners. This data set does not include all of the transaction relationships. This is the limitation of studies which use this data set. In this paper, we construct undirected network in which we do not make any difference between that firm A sells to firm B and that firm B sells to firm A. Because we believe that when analyzing clustering coefficient, it is more rational to study undirected network. Furthermore, not only when studying clustering coefficient, undirect network is fundamental than directed one. Thus the relation in this paper is that whether there is a transaction between two firms. We built adjacency matrix which is a common method in analyzing network. In adjacency matrix, we set element-ij 1 if there is any transaction irrespective of whether buy or sell and from which to which between firm-i and firm-j. We set element-ij 0 if there is not any transaction between firm-i and firm-j. Hence, the number of transaction between firms is not considered neither, only whether there exists any transaction matters. Actually, from the data we are not able to know the number of transactions between specific firms. Thus, the new data which includes the number of transactions must reveal something important.

2.2.2 How To Calculate Degree Distribution

Now, we present the way of calculating degree distribution in detail. The way is as follows. First, count the number of all of the links from the vertex with degree k. We let denote "All degree(k)" this. Remember that degree means how many links the vertex has. Second, calculate $P(k)$ by

$$P(k) = \frac{\text{All degree}(k)}{\sum_{k'} \text{All degree}(k')} \tag{5}$$

To detect scale free distribution, it is better way to draw CDF rather than PDF, detailed discussion of which is written in S.N.Dorogovtesev and J.F.F.Mendes

(2003). Thus we will illustrate CDF.

2.2.3 Clustering coefficient

We introduce the definition of clustering coefficient. The clustering coefficient is defined for each vertex. For example, the clustering coefficient of vertex-j is defined as,

$$C_j = \frac{\text{The number of triangles around vertex-j}}{\text{The maximum number of possible triangles around vertex-j}} \tag{6}$$

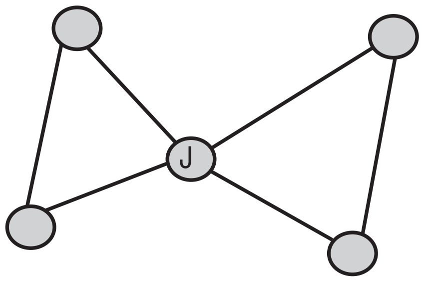

Figure 2: explanation for clustering coefficient

Fig.2 is the explanation of clustering coefficient. The number of triangles around vertex-j is 2, while the maximum number of triangles we can make around vertex-j is ${}_4C_2 = 6$, because there are 4 vertices around vertex-j. Thus, the clustering coefficient for vertex-j is $\frac{2}{6} = \frac{1}{3}$. If degree is one, we cannot define clustering coefficient. Because we cannot make any triangle around the vertex. Recall that the definition of degree is how many edges the vertex has. For example, in Fig.2, the degree of vertex-j is 4. Clustering coefficient is the method, in a nutshell, to measure how dense the vertices are connected locally among its neighbors. Thinking of friend network if the clustering coefficient is large, a friend of friend is also your friend. However if the clustering coefficient is small, a friend of friend is not your friend. It is worth noting that many real networks show high clustering coefficient than that of random network and small mean path length. Mean path length is defined as the average of path length over all pairs of the vertices.

2.3 Results and discussions

2.3.1 Degree Distribution of the Network

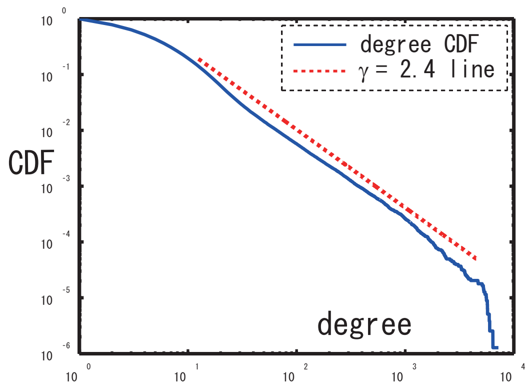

Figure 3: Degree distribution of Japanese firms' transaction network

We studied Japanese firms' transaction network.
Fig.3 illustrates $1 - \mathrm{CDF}$ of actual degree distribution in log-log plots by solid line. Saito et al. (2007) showed that directed network of Japanese firms' transaction in which the transaction between buy and sell were distinguished had scale free distribution. On the other hand, We show that degree distribution of undirected network[3] in which if there exists either transaction we regard the firms are linked follows scale free distribution such as $P(k) \sim k^{-2.4}$ also. $1 - \mathrm{CDF}$ of $P(k) \sim k^{-2.4}$ is illustrated in log-log plots by dashed line in the same figure. The drop of the actual data line in the right of the figure comes from the fact that there are only finite number of vertices in the network. If we want to see perfect scale free distribution, we need network with infinite vertices.

2.3.2 Hierarchical Structure of Japanese Firms

In this section we discuss hierarchical structure of Japanese firms' transaction network implied by clustering coefficient.

[3] In undirected network adjacency matrix is symmetric, while in directed network adjacency matrix is not generally symmetric.

Now, we demonstrate the clustering coefficient of 800,000 Japanese firms' transaction network. Fig.4 illustrates scatter plots and estimated line. x-axis is log(k), y-axis is log(clustering coefficient). Table 1 shows the estimation result.

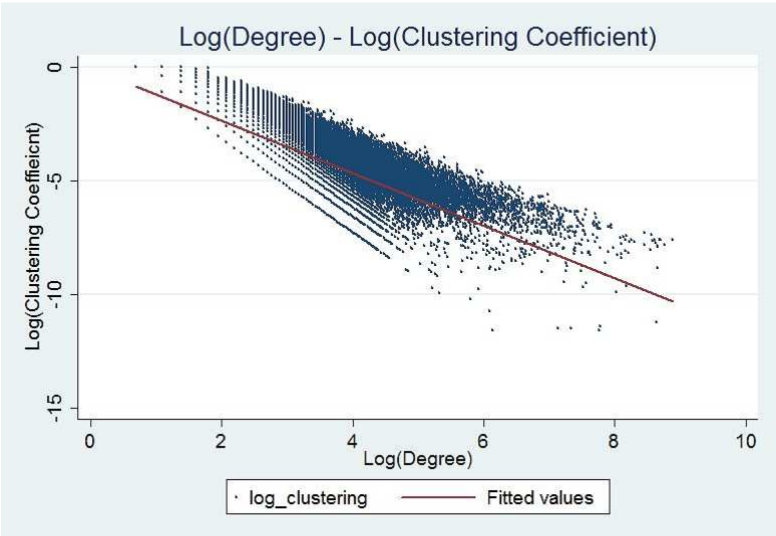

Figure 4: Log(Degree) - Log(Clustering Coefficient)

Table 1: Estimation results:Clustering coefficient

Variable	**Coefficient**	**(Std. Err.)**
log(degree)	-1.159	(0.002)
Intercept	-0.049	(0.004)

The estimated relation is,

$$\log C(k_j) = -1.159 \log k_j - 0.049 \tag{7}$$

R^2 of this estimation is 0.66. Remember that $C(k)$ stands for clustering coefficient of vertex whose degree is k. Eq.(7) strongly demonstrates that coefficient of log(k) is very close to -1 and this relation is equivalent to $C(k) \sim k^{-1}$, which is desired. As we discussed previously, this relation implies that Japanese firms' transaction network has not only scale free structure but also has hierarchical structure which is clearly exemplified (Ca) in Fig.1.

In Fig.4 there seems to be some structures of dots for Log(Degree) of less than 4 aligning on lines with negative slope. We need to explain why is this. Remember that clustering coefficient is defined as

$$
\begin{aligned}
C(k) &\equiv \frac{\text{The number of triangles}}{k(k-1)/2} \\
&\sim \frac{\text{The number of triangles}}{k^2/2}
\end{aligned}
\tag{8}
$$

However, the number of triangles in eq.(8) is discrete such as 1, 2, 3 and so on. Hence, the bottom structure consists of the points the number of triangles of which is 1 then the clustering coefficient is $2 \times 1/k^2$. Similarly, the second bottom structure consists of the points the number of triangles of which is 2, subsequently the clustering coefficient is $2 \times 2/k^2$. The clustering coefficient of other structures are $2 \times 3/k^2$ and so on. Since we take logarithm of them, the slope of these structures are -2, so that where the number of triangles is small they seem align.

We would like to mention followings. Barabasi and Albert (1999) introduced a famous mechanism generating scale free network known as preferential linking. In a nutshell, the more degrees the vertex has the more links it attracts from other vertices. However, it is well known that the network generated by this preferential linking mechanism does not have hierarchical structure. In this network, the relation $C(k) \sim k^{-1}$ cannot be observed. Because transaction network has hierarchical structure, there must be another mechanism which makes firms' transaction network.

2.3.3 Degree Correlation

Another important discovery is the existence of degree-degree correlation.

In the network degree and next neighbor degree has the following relation.

$$k_{nn} \sim k^{-0.5} \tag{9}$$

k_{nn} stands for next neighbor degree.

The result is $k_{nn} = 1289\ k^{-0.546}$. Table 2 shows regression results. This shows $\log(k_{nn}) = -0.546\ \log(k) + 7.162$, which is almost same as eq.(9). R^2 of this regression is 0.681.

Fig.5 illustrates scatter plot and fitted curve, Fig.6 demonstrates in log log plot and fitted curve.

Table 2: Estimation results: Degree Correlation

Variable	Coefficient	(Std. Err.)
log degree	-0.546	(0.014)
Intercept	7.162	(0.080)

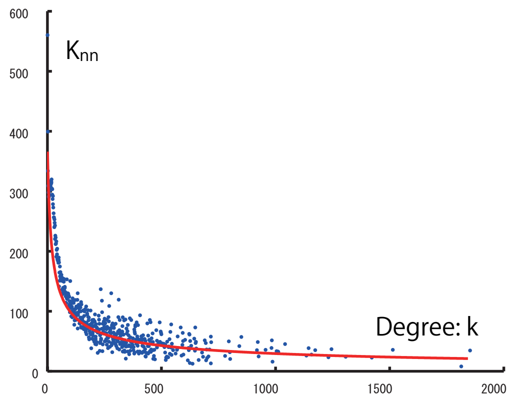

Figure 5: Degree Correlation

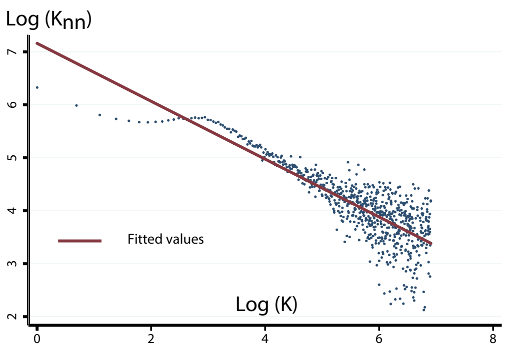

Figure 6: In Log Log Plots

3 Conclusion

We studied Japanese firms' transaction network by analyzing degree distribution, clustering coefficient and degree correlation. Subsequently, we discovered following three important properties.

First, we found undirected network is scale free network. Second, We discovered the network has hierarchical structure. Third, discovery of degree correlation.

As mentioned earlier, we believe that the study of actual network will lead to further research which reveals hidden relation between network structure and economy. We need information as to actual network structure when we build models on network. In many other areas of science, it was discovered that micro and macro properties are dependent on network structure. It depends on, for example, whether random network or scale free network, clustering coefficient and so on. We expect that similar relation will be discovered in Economics as well. In this sense, to study actual network and to provide information is very important.

References

Abramson, G. and Kuperman, M. 2001, Social games in a social network, *Physical Review E*, **6303** (3), p. 4.

Albert, R. and Barabási, A.-L. 2002, Statistical mechanics of complex networks, *Rev. Mod. Phys.*, **74** (1), pp. 47–97, Jan.

Aoki, M. and Yoshikawa, H. 2006, *Reconstructing Macroeconomics: A Perspective from Statistical Physics and Combinatorial Stochastic Processes*: Cambridge University Press.

Barabasi, A.-L. and Albert, R. 1999, Emergence of scaling in random networks, *Science*, **286** (5439), pp. 509–512.

Barabasi, A.-L. and Oltvai, Z. N. 2004, Network biology: understanding the cell's functional organization, *Nature Reviews Genetics*, **5** (2), pp. 101–113. 10.1038/nrg1272.

di Bernardo, M., Garofalo, F., and Sorrentino, F. 2005, Synchronizability of degree correlated networks, *Arxiv preprint cond-mat/0504335*.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D. U. 2006, Complex networks: Structure and dynamics, *Physics Reports*, **424** (4-5), pp. 175–308.

Callaway, D., Hopcroft, J., Kleinberg, J., Newman, M., and Strogatz, S. 2001, Are randomly grown graphs really random?, *Physical Review E*, **64** (4), p. 41902.

Di Bernardo, M., Garofalo, F., and Sorrentino, F. 2007, Effects of degree correlation on the synchronization of networks of oscillators, *International Journal of Bifurcation and Chaos*, **17** (10), pp. 3499–3506.

D.J.Watts 1999, *Small World:The dynamics of Networks between Order and Randomness*: Princeton University Press.

Dorogovtsev, S. N. and Mendes, J. F. F. 2002, Evolution of networks, *Advances in Physics*, **51** (4), pp. 1079 – 1187.

Garlaschelli, D., Battiston, S., Castri, M., Servedio, V., and Caldarelli, G. 2005, The scale free nature of market investment network, *Physica A*, **350** (2-4), pp. 0310503–491.

H.Strogatz, S. 2001, Exploring complex networks, *Nature*, **410**, p. 268.

Jeromos, V. and Gyorgy, S. 2005, Evolutionary prisoner's dilemma game on hierarchical lattices, *Physical Review E*, **71** (3), p. 036133.

Motter, A., Zhou, C., and Kurths, J. 2005, Enhancing complex-network synchronization, *Europhysics Letters*, **69** (3), pp. 334–340.

Newman, M. 2002, Assortative Mixing in Networks, *Physical Review Letters*, **89** (20), p. 208701.

Newman, M. E. J. 2003, The structure and function of complex networks, *SIAM Review*, **45** (2), pp. 167–256.

Ohtsuki, H., Hauert, C., Lieberman, E., and Nowak, M. A. 2006, A simple rule for the evolution of cooperation on graphs and social networks, *Nature*, **441** (7092), pp. 502–505.

P.Erdos and A.Renyi 1959, On random graphs, *Publicationes Mathematicae*, **6**, pp. 290–297.

P.Pastor-Satorras and A.Vespignani 2004, *Evolution and Structure of the Internet: A Statistical Physics Approach*: Cambridge University Press.

Ravasz and Barabasi 2003, Hierarchical organization in complex networks, *Physical Review E*, **67** (2), p. 026112.

Saito, Y. U., Watanabe, T., and Iwamura, M. 2007, Do larger firms have more interfirm relationships? *Physica A*, **383** (1), pp. 158–163.

Santos, F. C. and Pacheco, J. M. 2005, Scale-free networks provide a unifying framework for the emergence of cooperation, *Physical Review Letters*, **95** (9), p. 098104.

S.Bornholdt 2003, *Handbook of Graphs and Networks: From the Genome to the Internet*: Wiley-VCH.

S.N.Dorogovtesev and J.F.F.Mendes 2003, *Evolution of Networks*: Oxford University Press.

Sorrentino, F., di Bernardo, M., Cuellar, G., and Boccaletti, S. 2006, Synchronization in weighted scale-free networks with degree–degree correlation, *Physica D: Nonlinear Phenomena*, **224** (1-2), pp. 123–129.

Souma, W., Fujiwara, Y., and Aoyama, H. 2006, Change of ownership networks in Japan, in *Practical Fruits of Econophysics: Proceedings of the Third Nikkei Econophysics Symposium*, Springer Verlag.

Vega-Redondo, F. 2007, *Complex Social Networks*: Cambridge Univerity Press.