Proceedings of the 8th Polish Symposium of Physics in Economy and Social Sciences FENS, Rzeszów, November 4–6, 2015

Identification of Insider Trading Using Network Numerical Models

A. Jakimowicz and A. Baklarz

Institute of Economics, Polish Academy of Sciences, Palace of Culture and Science,
pl. Defilad 1, PL-00901 Warsaw, Poland

This article presents a network algorithm for identifying transactions which may constitute a violation of restricted periods, namely, making transactions in company shares by persons possessing inside information. The empirical research was performed on the basis of publicly available information on exchange trading, originating from the Warsaw Stock Exchange. The analysis is based on a numerical model which describes information spreading in a network with an information bottleneck. The applied method can confirm with high probability the use of inside information for carrying out unauthorized stock market transactions.

DOI: 10.12693/APhysPolA.129.980

PACS/topics: 89.65.Gh, 89.65.–s, 05.65.+b, 89.75.–k

1. Introduction

The issue of information asymmetry has been very extensively discussed in the literature in view of its particular importance. In the case of information asymmetry, effective allocation of resources is not possible, which may lead to inefficiency of market mechanisms. The issue of information asymmetry also occurs in cases of adverse selection and moral hazard [1, 2].

Information asymmetry is an issue of particular importance for capital markets. Since this phenomenon can, to a greater or lesser extent, disturb operations of capital markets, it is governed by relevant legal regulations [3]. Consequently, it is reasonable to claim that the effect of information asymmetry on prices is of high importance for research on financial markets, which has resulted in — among others — developing the information market hypothesis [4–7].

The subject matter of information asymmetry, due to its key importance for capital markets, has been reflected in legal regulations concerning trading in financial instruments in many countries. In Poland, the basic legal act in this field is the Act on Trading in Financial Instruments of 29 July 2005 (Dz.U. 2005, No. 183 item 1538 as amended), which specifies in its Art. 156, a catalogue of persons who are not allowed to use inside information. This catalogue, starting with the phrase "Whoever", after listing specific persons, ends with a clear statement: "may not use such information". Actually, this catalogue lists all persons who could obtain inside information and have knowledge about the confidential nature of this information.

Additionally, the Act on Trading in Financial Instruments, in Art. 159.1 and Art. 159.1a clearly specifies behaviour prohibited for persons who could have — due to their functions held — wider access to inside information. In Sect. 2 of this article, the Act defines a restricted period, i.e. the period in which inside information cannot be used. In case of financial statements, these are the following periods: two months before releasing the annual statement, one month before releasing the semiannual statement and two weeks in case of the quarterly statement. For other information, the restricted period runs from the moment of obtaining the information until its disclosure by the company.

Compliance with the Act is supervised by the Polish Financial Supervision Authority. Pursuant to Art. 174 of the Act on Trading in Financial Instruments, the Authority may impose a fine up to PLN 200,000 (USD 50,000) in case of breaching Art. 159. Until now, eleven persons have been punished under this Article. The last penalty was imposed in 2012. Penalties were imposed on five persons holding positions in management boards of the companies, one person who was the main company shareholder and employee, three "insiders" and two natural persons with an undefined relationship towards the company. In total, the fines amounted to PLN 347,000 (USD 86,750) [8].

It is highly unlikely that, despite such severe sanctions, the phenomenon of using inside information disappeared in 2012. It should be rather presumed that the nature of those abuses has changed. It is probable that multiple small transactions are carried out by persons closely related to persons with access to inside information. Due to their small scale, such transactions do not significantly affect the price of financial instruments. A change in price occurs some time later, namely, after the information is disclosed by the company.

The aim of this article is to present the potential of a network algorithm designed for identifying situations in which inside information could have been disclosed to external persons. This algorithm uses generally available data concerning share listings on the Warsaw Stock Exchange and official announcements published by companies. The basis of the algorithm is a numerical model describing the spreading of bottleneck information. The algorithm can be used by institutions supervising a capital market to identify possible infringements.

The rest of the paper proceeds as follows. Section 2 describes the network model for bottleneck information spreading. In Sect. 3, we present a pilot study for COMP SA, an IT company participating in the sWIG80 index. In Sect. 4, we show an extension of the research on subsequent six companies listed in the same index. Section 5 concludes.

2. The model for bottleneck information spreading

The legal regulations presented in Sect. 1 indicate that inside information, often carrying significant importance for investors and shareholders, cannot be made public before the expiry of restricted periods. However, as it has been demonstrated above, sometimes such information is transferred before its official release to a group of potential or current shareholders. The spreading of inside information can be described with the use of the information flow bottleneck model.

According to "small worlds" models, taking into consideration the well-known Milgram experiment and the Watts–Strogatz model with later modifications [9–12], if there only exist remote nodes in the network which are connected to each other, the path of information and its flow time is significantly reduced. Those models are used, e.g. for simulating spreading of epidemics and in simulations of information flow bottlenecks. In order to simulate bottleneck information spreading in a stock exchange market, a modified CODA (agents with continuous opinions and discrete actions) model was used [13].

In the CODA model, a change in the opinion (continuous variable) of nodes is revealed by observing the decisions (discrete variable) of neighbouring nodes. A decision taken by the neighbouring node can only have two discrete states: +1 or –1. Nevertheless, the premises are weighed continuously — the node verifies to what extent the given information is true and depending on this, it makes a binary decision. The version of the model useful for examining financial markets was developed on the basis of the innovation diffusion model, where two-dimensional networks are considered, made of 32×32 or 64×64 nodes, with equal distribution of links, with each node having four links. A CODA numerical model, adjusted for examining bottleneck information spreading, is based on the following assumptions:

a) There are two two-dimensional networks:
   - An internal network (from 2×2 nodes to 16×16 nodes),
   - An external network (from 30×30 nodes to 16×16 nodes).

b) There is one link between the networks (bottleneck — a blockage against using the information), i.e. one node in the internal network has five edges and one node in the external network has five edges.

c) The node can have a "binary status"; it either has information or does not have it (there are no intermediary states).

d) All nodes are tested in unit time.

e) The parameter is the probability of passing the information forward.

It should be emphasized that the division of the network into its internal and external parts can differ depending on the type of information. Therefore, the following cases may occur:

- For contract negotiations, the internal network may be small and include only 4–5 persons, while the external network can be relatively large, due to the pressure exerted on the competition and the contractor, i.e. can consist of about 100 persons;
- While drafting financial statements, the internal network can be relatively large, e.g. may consist of about 40–50 persons involved in preparing the statements, while the external network can be relatively small and include only several persons highly trusted in relation to the internal network.

Consequently, it is necessary to analyse the effect of both external and the internal networks on the emergence of inside information within the external network.

A selected course of simulation for the following case is presented below: internal network 5×5, external network 27×27 and probability of passing forward of 15% (a longer retention of information is a random factor).

Figure 1 presents the total number of persons knowing the information. As it shows, before the information was passed from the internal network to the external network, it was learnt by all persons in the internal network (which is consistent with intuition). It was only after a certain time that one of the persons decided to pass the information to the external network. Although the internal network could be described using the star model, it should be noted that the information flow in the internal network is not directly observed on the stock exchange by investors, therefore it is of no importance for spreading the inside information outside. Since the information flow is not investigated in the internal network, modelling of this network does not have to be limited to the star model.

Figure 1 does not specify the time unit. It can be a minute, an hour or even a day. The time unit is related to the issue of determining the probability of an individual transfer of information between the nodes. If this probability is determined at the level of 15% in one conversation, then the time unit is the time during which the nodes are in direct contact. Those time brackets should not be confused with restricted periods. The date of releasing inside information to the public is defined in the financial report schedule and specified by legal regulations. On the other hand, the above simulation shows

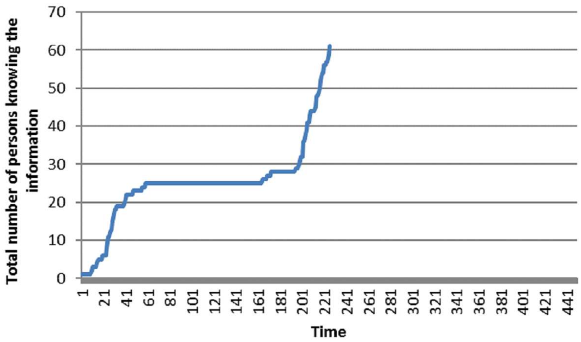

Fig. 1. The number of persons who have obtained the information in the bottleneck network. The internal network is 5 × 5, the external network — 27 × 27, and the probability of passing the information forward from one node to another is 15%.

how inside information can be spread in two groups of persons connected by only one link of a low probability of information flow. Therefore, the first time unit is not related to the other time units.

The height of the external part (the number of persons who have learnt the information) is determined by the network of a specific person holding the inside information, who passed it outside. It is also worth noting that after passing the bottleneck, the information spreads in the second network extremely fast and causes a significant reaction. What is important, the model examines the reaction of individual persons (private investors) and not institutions.

It should be emphasized that the model is based on the assumption that only one person from the internal network will pass the information forward, with a relatively low probability. A modification of the model, consisting in taking into account a larger number of persons who will pass the inside information forward, will not change the characteristic shape of the curve presented in Fig. 1, but the time in which the information passes to the other network will be reduced.

The total probability of the leak P_{total} can be calculated from the following formula:

$$P_{\text{total}} = 1 - (1 - p_1) \times (1 - p_2) \times \ldots \times (1 - p_n), \quad (1)$$

where symbols $p_1, \ldots, p_n$ denote the probability of passing information forward by individual persons. From formula (1), it can be calculated that for the above specified internal network ($n = 5 \times 5 = 25$ persons) with a unit probability of the leak at level $p_1 = p_2 = \ldots = p_n = 15\%$, the total probability is 98%. Therefore, an event consisting in the occurrence of the leak is almost certain. In other words, the total probability of a leak amounting to 98% shows that even if a unit probability of the leak is low (which is ensured by high financial penalties for such an offence), it is practically certain that inside information will reach the external network.

In order to determine whether insider information has been disclosed, it should be specified whether transactions occurring in predictable periods consistent with periods specified in the Act on Trading in Financial Instruments could affect the above-average numbers of transactions in company shares in short time ranges. Therefore, it can be assumed that the number of transactions in given shares, incrementally calculated at the stock exchange, in the period equal to the restricted period applicable for financial statements and the period of up to 30 days for other transactions, will have a shape similar to the shape presented in Fig. 1.

3. A pilot study for COMP SA

A preliminary study was carried out for COMP SA, an IT company, i.e. a company with good access to IT tools facilitating information circulation. The company has a diversified shareholder structure: 49.5% shares are held by pension and investment funds, 14% shares are held by the President of the Company, 14% by subsidiaries, and other shareholders hold 22.5%. In 2012, COMP reported a significant loss, and in 2013, it disclosed a high profit along with increased turnover. The listing data cover the period between 1 January 2010 and 31 December 2014, i.e. 1,248 days of transactions in total. The data originate from the archive of the Warsaw Stock Exchange listings [14]. The persons with access to inside information (in particular to financial information) can be classified as follows:

- Management Board: 5 persons.
- Supervisory Board: 7 persons.
- Audit Committee of the Supervisory Board: 3 persons.
- Advisers to the President: 6 persons.
- Management boards of the subsidiaries: 19 persons.
- Accounting department employees: more than 10 persons.

The list includes all persons related to the company who cannot reveal or use inside information. This prohibition concerns all shareholders, regardless of the level of their concentration of capital and this is what is meant as avoidance of information asymmetry economics. It should be emphasized that for fiscal reasons, financial statements of the company (in particular annual and semi-annual statements) are published with a significant delay (over one month).

In order to develop an algorithm for identifying days of transaction with atypical behaviours, the following steps were taken:

1. Two separate time series were constructed:
   a) Days on which the average value of a unit transaction exceeded the median. The average number of transactions for higher value

transactions is 15 (i.e. "fat" right tail). The distribution of the number of transactions is normal, with probability of more than 99% (Fig. 2).

b) Days on which the average value of a unit transaction was lower than the median. The average number of transactions for low values transactions is 7. The distribution of the number of these transactions is also normal, with probability of more than 99%, although it includes significant days with a high number of transactions (Fig. 3).

2. The number of transactions was cumulatively calculated for each time series (9,695 transactions for large transaction days and 4,387 transactions for days of transactions below the median). The charts were standardized to 1. It consisted in checking the value of the median of transactions for the previous 365 calendar days. Therefore, it was calculated as the value of 12 months preceding the day of dividing transactions into large and small. The calculated value of this median was compared to the value of an average transaction (i.e. a quotient of the turnover for a given day and the number of transactions on a given day). If the median was higher, then a given day was included into the time series of small transactions. Otherwise, it was classified as a large transaction. The above data were derived from the total number of transactions in both time series.

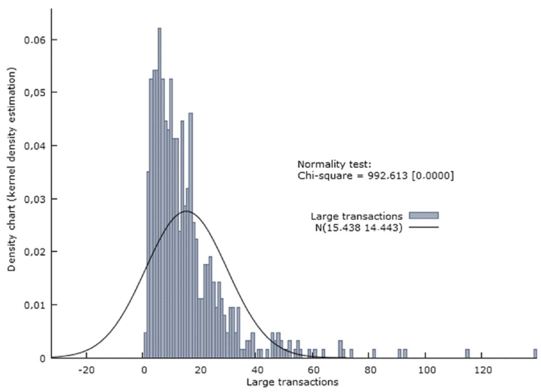

Fig. 2. Distribution of transaction density for transactions above the median value.

3. The numerical model assumes that a significant leap in the number of transactions must take place within maximum two days as of the leak. As the information flow bottleneck model (Sect. 2) shows, after the information is transferred from the internal to the external network, it is later spread immediately, as a matter of principle. For stock exchange listings, two days make a good reflection of the term "immediately".

4. The time series of transactions above the median was treated as a reference series. The time series of transactions below the median was examined for a leap occurring within two days, of the value of at least 5 times of the standardized time series of large transactions. This is a numerically selected value that describes the size of the external network, and in this case corresponds to at least 35 additional transactions a day. According to Fig. 3, it is responsible for the other distribution ridge, with the centre between 25 and 40 transactions. This ridge was taken into account because of experimentally selected factor 5, which is explained below. Consequently, the points satisfying the following inequality were searched for:

$$\frac{\sum_{i=1}^{t}\text{No. of small tr}_i - \sum_{i=1}^{t-2}\text{No. of small tr}_i}{\sum_{i=1}^{1248}\text{No. of small tr}_i[t-(t-2)]} >$$

$$\alpha\frac{\sum_{i=1}^{t}\text{No. of large tr}_i - \sum_{i=1}^{t-20}\text{No. of large tr}_i}{\sum_{i=1}^{1248}\text{No. of large tr}_i[t-(t-20)]}, \quad (2)$$

where $\alpha = 5$, t is the measurement point, $t-2$ is a point 2 listings earlier, while $t-20$ is a point 20 listings earlier. Value $\alpha = 5$ was determined experimentally and the research on the sensitivity of results to this parameter is still ongoing. The $t-2$ lag was assumed according to previous explanations (in point 3), while the $t-20$ lag refers to the Fama semi-strong financial market efficiency hypothesis [4]. Significant announcements provided by the company to the Warsaw Stock Exchange were also assigned to transaction days. There were 119 such days in the entire set.

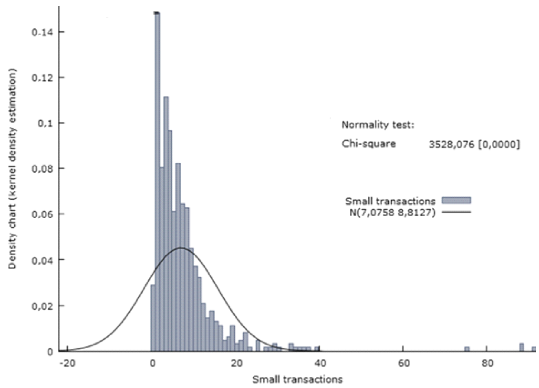

Fig. 3. Distribution of transaction density for transactions below the median value.

The research assumed, according to the efficient market hypothesis, that in the period of 12 months, the mean number of transactions has a Gaussian distribution with 0 as the cut-off point.

It was necessary to standardize the transaction numbers in view of the fact that the number of transactions for large transaction time series differed from the num-

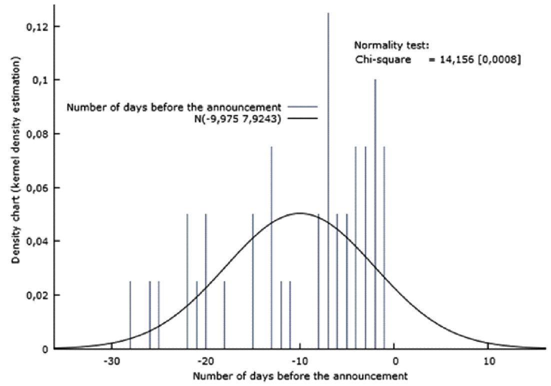

Fig. 4. Distribution of density of the number of advance days for a transaction leap before the announcement for transactions below the median value.

ber of transactions for a small transaction time series. If both time series were affected only by the same factors, i.e. external information, then both series would have similar behaviour and inclination. The occurrence of a fast, i.e. occurring within 2 days, jump in the number of transactions without a noticeable jump in the inclination on the large transaction time series reacting in a different way may indicate an illegal use of inside information.

An additional 35 transactions, i.e. five times the normal value (7 transactions daily), occurring on one day indicates the existence of atypical market phenomena, which are not simultaneously reflected in the time series of large transactions. This lends credence to the fact of transferring information which is not yet known to all investors (and, in particular, is not known to large investors) to the external network of chosen investors.

Changes in share prices have been omitted in the study due to the fact that they can have multiple reasons, not always related to the leak of inside information (e.g. general information on the condition of financial markets in the world, inflow of foreign investors as an effect of a good rating of Poland, etc.). For this reason, the division into 2 time series was introduced so that the series of large transactions could be used as a referential series in relation to the examined time series of small transactions. In other words, as a result of this division and of focusing on the analysis of small transaction time series, an analysis of changes in share prices was deliberately avoided. Although the time series for COMP was previously analysed by the authors (in an unpublished study) from the perspective of the semi-strong financial market efficiency hypothesis (i.e. in terms of the rate of price change after receiving information), the existence of a semi-strong form of efficiency was neither confirmed nor denied for this company.

After carrying out tests, 40 leaps in the number of transactions were found, for which the average advance period in relation to the announcement was almost 10 days. The leaps preceding the announcements also included days with the largest number of small transactions (88 transactions — 3 days of the advance and 91 transactions — 20 days of the advance). The distribution of advance days demonstrates relatively short advance periods — usually between 1 and 7 days (Fig. 4).

4. Extending the research on subsequent companies

In order to establish whether the issue of an inside information leak is common or incidental, the algorithm for identifying atypical transactions was verified, by examining six companies participating in the sWIG80 index (companies with numbers 29, 34, 38, 45, 48 and 51 in the sWIG80 index) in the same way. During the verification of the algorithm, parameters set for the analysis of COMP SA data were not changed (Table I). The algorithm presented in Sect. 2 was applied.

It is worth noting that such research is continued and that it includes a much higher number of companies. However, the research has not yet been published. Additionally, studies are being conducted on information other than inside information, which may affect the number of transactions. In particular, the analyses concern the effect of discussions held on the Internet forums for individual investors on the number of transactions on the Warsaw Stock Exchange, but that research has not revealed any effects of other information available to small investors on the numbers of transaction.

TABLE I

Results of verifying transactions below the median in terms of the advance period in relation to the announcement made by the company. 1 — number of significant announcements, 2 — number of points with suspicious correlation to subsequent announcements, 3 — average number of days of the advance, 4 — median of the number of days of the advance, 5 — percentage of leaps in relation to announcements [%]

company	1	2	3	4	5
COMP	119	40	9.68	7	33.6
29_Lentex	289	19	9.89	10	6.5
34_Robyg	129	26	10.77	8.5	20.2
38_Impexmet	74	33	13.44	13	44.6
45_Kopex	422	12	10.25	5.5	2.8
48_ZPUE	118	42	14.70	14	35.6
51_Polnord	183	53	9.98	7.5	28.9
TOTAL	**1334**	**225**	–	–	**16.9**

The knowledge of significant information was obtained in the following manner. Communications provided by companies to the Stock Exchange were assigned to individual days of transactions. All communications were analysed, and the ones with lower importance assigned were eliminated. The communications in which the following phrases or fragments of words occurred were not taken into consideration: “przetar”, “skład”, “wykaz”, “wybór”, “rad”, “podjęt”, “chce”, “moż”, “opinii”, “opis”, “wywiad”, “przekazy”, “sprostow”. Additionally, the communications were analysed for each company individually and the content specific for a given company was added to the rules of elimination. The analysis of communications was carried out by Baklarz, a chartered auditor, on the basis of expert knowledge.

The examination of subsequent companies with the use of this method led to the conclusion that there are on average four points, of various distribution of the number of advance days in relation to the announcement, per company.

Probably the most frequent reason for an increase in the number of transactions before the announcement was the expected financial result. Other potential reasons for increasing the number of transactions before the announcement included:

- A resolution on the issue of shares or bonds;
- Concluding a significant agreement;
- Starting a winding-up procedure for a subsidiary;
- Renewal of a credit line;
- Annex to a significant agreement.

In case of unexpected information and quarterly reports, the time between the leap in the number of transactions and publication of the announcement was clearly shortened to a few days (between 1 and 5), while for semi-annual or annual reports it was clearly extended to even more than 30 days.

5. Conclusions

The algorithm presented in this article, used for identifying transactions for which Art. 159 of the Act on Trading in Financial Instruments was probably violated, is based purely on publicly available data. Therefore, it can be only statistically claimed that a probability of a certain regularity exists. At the same time, it should be noted that the phenomenon of passing the information through the "bottleneck" for legal reasons is generally very unlikely. Nevertheless, as results from the numerical models, even at the probability of 15% and an adequately long time period before the official disclosure of information, there is a high chance that the inside information can reach the network of trusted persons. The analysed examples revealed 225 such cases (Table I). Obviously, it cannot be definitely established on the basis of this algorithm whether the information leak really occurred or it was just a coincidence. However, as indicated by the distribution of the number of transactions, the probability of an error is rather low in case of low value transactions. Further improvement of the algorithm should be oriented towards testing sensitivity to individual keywords and examining the probability of a unit leak.

The developed algorithm is a useful tool aimed at preliminary selection of suspicious transactions recorded on the stock exchange. Its application by financial supervision authorities could lead to eliminating unauthorized use of inside information in transactions. However, definite establishment of the fact of using confidential data is possible only when the actual data concerning specific transactions and mutual networks of relations are available. A great analogy here is the analysis of behaviours in criminal groups, using the COPLINK system as an example, which in 2002 included over 30,000 nodes (persons), between whom more than 80,000 links existed [15].

References

[1] G.A. Akerlof, *Quart. J. Econ.* **84**, 488 (1970).

[2] B. Holmström, *Bell J. Econ.* **10**, 74 (1979).

[3] F.J. Fabozzi, F. Modigliani, M.G. Ferri, *Foundations of Financial Markets and Institutions*, Prentice Hall, Englewood Cliffs 1994.

[4] E.F. Fama, *J. Fin.* **25**, 383 (1970).

[5] S.J. Grossman, J.E. Stiglitz, *Am. Econ. Rev.* **70**, 393 (1980).

[6] R.J. Shiller, *Am. Econ. Rev.* **71**, 421 (1981).

[7] G.D. Saxton, A.E. Anker, *J. Commun.* **63**, 1054 (2013).

[8] Polish Financial Supervision Authority, *Kary nałożone przez KNF*, November 11 (2014) (in Polish).

[9] S. Milgram, *Psychol. Today* **1**, 61 (1967).

[10] J. Travers, S. Milgram, *Sociometry* **32**, 425 (1969).

[11] D.J. Watts, S.H. Strogatz, *Nature* **393**, 440 (1998).

[12] A. Fronczak, P. Fronczak, *The World of Complex Networks. From Physics to the Internet*, Wydawnictwo Naukowe PWN, Warszawa 2009 (in Polish).

[13] A.C.R. Martins, C. de B. Pereira, R. Vicente, *Physica A* **388**, 3225 (2009).

[14] GPWInfoStrefa, archiwum notowań, (in Polish).

[15] H. Chen, in: *Digital Libraries: People, Knowledge, and Technology*, Eds. E.-P. Lim, S. Foo, C. Khoo, H. Chen, E. Fox, S. Urs, C. Thanos, Springer-Verlag, Berlin 2002, p. 36.