Detecting Wash Trade in the Financial Market

Yi Cao, Yuhua Li, Sonya Coleman, Ammar Belatreche, T.M.McGinnity
Intelligent Systems Research Centre
University of Ulster
Londonderry, United Kingdom
cao-y1@email.ulster.ac.uk, {y.li, sa.coleman, a.belatreche, tm.mcginnity}@ulster.ac.uk

Abstract—**Wash trade refers to the activities of traders who utilise deliberately designed collusive transactions to increase the trading volumes for creating active market impression. Wash trade can be damaging to the proper functioning and integrity of capital markets. Existing work focuses on collusive clique detections based on certain assumptions of trading behaviours. Effective approaches for analysing and detecting wash trade in a real-life market have yet to be developed. This paper proposes a new analysis approach for abstracting the basic structures of wash trade based on the network topology theory and a novel approach for detecting wash trade activities. The evaluation experiments conducted on four NASDAQ stocks suggest that wash trade actions can be effectively identified based on the proposed algorithm.**

I. INTRODUCTION

Surveillance of financial exchange markets for preventing market abuse activities has been attracting more attention from academic and industry since the financial crisis in 2008 and especially since the flash crash in 2010. One major form of market abuse is wash trade, which occurs when the same individuals or a group of collusive clients are on both the selling and buying sides of a transaction, therefore there is no beneficial change in ownership of the traded financial instruments. By the non-bona fide collusive transactions, wash trade has the effect of creating a misleading appearance of an active interest in the stock [1]. Most of the existing related literature empirically studies the collusive cliques based on some certain assumptions of the "activity similarity". Very few studies address the analysis of the features of different wash trade scenarios and the corresponding detection approaches.

To conduct a wash trade, traders usually set up an agreement to construct a network and then trade following the connection. The network can simply be comprised of only one or two traders and can also be comprised of a lot of traders connecting in complex hierarchies. Therefore, discovering the collusive network is crucial for detecting wash trade activities.

We propose a novel detection approach that considers a complete spectrum of the problem. In this approach, the collusive structure of a wash trade is analysed and summarised based on the network topology theory. A recursive algorithm is then proposed based on the analysis of the network topology. To apply the algorithm, the trading records are aggregated and then sorted according to the traded volumes. Substantial experiments are conducted on real data from primary NASDAQ market for testing the practicability in real-life. The obtained results show the effectiveness of the proposed

algorithm on discovering the wash trade activities.

The reminder of the paper is organised as follows: Section II provides a brief review of wash trade manipulation and the corresponding detection methods. The structure of the wash trade transactions is analysed in Section III as well as the detection algorithm and the pre-organization scheme. Experimental evaluation of the proposed approach is provided in Section IV. Finally, Section V concludes the paper and discusses potential improvements and future work.

II. RELATED LITERATURE

Wash trade intends to increase market transaction volume, with the aim of giving the impression of active stock activities. The UK Financial Services Authority (FSA) defines wash trade as a sale or purchase action where there is no change in beneficial interest or market risk, or where the transfer of beneficial interest or market risk is only between parties acting in concert or in collusion, other than for legitimate reasons [2]. The Committee of European Securities Regulators (CESR) further indicates that a wash trade is the deliberate arrangement in concert or collusion [3].

To the best of our knowledge, there is no reported work on the detection of wash trade activities in capital markets. The only research that comes close is the work on the detection of collusive cliques based on similar trading behaviours, which are defined as the members of cliques buying (or selling) certain equities in a similar way. A spectral clustering based approach was developed [4], where a trading-behaviours network is generated and any behaviour that deviates from the network is reported as irregularities. The assumption of this work is the strong consistence of a trader's current behaviours and his/her previous trading network. A graph clustering algorithm for detecting a set of collusive traders has been proposed [5]. The relationship between traders is constructed as a stock flow graph, and those with "heavy trading" within their network are clustered as collusion set. A new approach to the correlation matrix of one trading day was presented in recent work [6], where trader behaviours are represented by an "aggregated time series of signed volumes" of submitted orders. The similarities in behaviour between two or more traders are measured using Pearson's product-moment correlation coefficient. If the correlation coefficient is larger than a user-specified threshold, the corresponding traders are considered a potential collusive clique.

To date, academic research has mainly focused on the detection of trading collusions according to analogous trading behaviours [5] [6], which were defined by aggregated order sequences across various stocks. Detecting the overall behaviours across different stocks can hardly reach a

conclusion with a precise and determinable manipulation activity on a certain instrument. Thus the detected collusions only show a collective correlation between their trading sequences rather than a particular wash trade action. In the meantime, the regulations [2] [3] in recent years broadened the gap between the requirements and the lack of an effective wash trade detection tool. However, there is no reported work on the analysis of wash trade behaviour or a detection approach covering a complete spectrum of wash trade tactics. Given the gap in the field, it is this aspect of market manipulation that the paper seeks to address.

III. WASH TRADE AND ITS DETECTION

As the FSA and CESR pointed out in their consultation reports [2] [3] that it is hard to distinguish a wash trade because the format of trading collusions varies and the collusive transactions can be buried in mass numbers of normal trading records. The basic structure of collusive transactions is theoretically defined as a single trader simultaneously buying and selling [2] [3] and sale (or purchase) of blocks of shares through multiple traders as passing the parcel between them [7]. In practice, the collusive trading can be more complex than the basic two definitions, e.g., a complex mesh transaction network reported by Nanex on 31 May 2013 [8]. The lack of accurate formulation of the collusive transaction network led to the scarcity of effective wash trade detection tools.

Wash trades can be carried out by either one single trader or multiple traders [5] [6] [7]. The arrangement of one or various traders in a trading collusion is essentially analogous to a logically topological structure of various nodes in a network [9]. The analogy can be summarized as:

- The network topology is comprised of a number of nodes, which pass data to each other along the connections. The trading collusion contains numbers of traders, who buy or sell certain stocks following the specific connection agreement.

- The network in reality is complex and is usually composed of several basic network topologies [9]. The basic trading collusion is defined while the collusive transactions in the market are more complicated and might be explained as the hybrid of the basic formats.

Therefore, the transactions among traders within a collusion network can be analysed by the network topology theory. The analysis aims to identify the most basic structures that comprise the complex network and build the detection approach according to the basic structure.

A study of network topology identifies five most common topologies, bus, ring, star, tree and mesh [10]. A trading transaction occurs between a seller and a buyer (point-to-point connection in a topology). The bus topology, where each node connects to a shared bus instead of another node, does not fulfil the trading context. Therefore, we analyse the wash trade scenario by a collusive network of four different topologies: ring, star, tree and mesh.

A. Wash trade topologies

1) Ring topology

Ring topology with only one node is the simplest network. It shows the basic wash trade format: one single trader "A" simultaneously buying and selling identical shares of a stock [2] [3] as illustrated in Fig. 1.
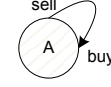

Fig. 1 Wash Trade - single trader.

Additional to the basic "single-trader" format, a wash trade can also be carried out among a network of N traders. Let's set $N = 7$ as the example in our following analysis.
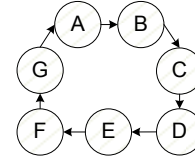

Fig. 2 Ring topology of seven traders.

In a ring topology wash trade scenario with multiple traders, every trader connects to exactly two other trading neighbours forming a single continuous pathway for transactions through each trader in one single direction (either "clockwise" or "counter clockwise") with each trader along the way handling every selling and buying action. This is defined as "passing the parcel" in [7]. After a complete transaction loop, the beneficial interest has been transferred across the collusive group and no traders in the collusion have an actual position change. An example of wash trade transactions following the ring topology in Fig. 2 is illustrated in Table 1, where Δv is defined as a small positive number of volume. The traded volumes between any two neighbours are not necessarily 100% identical since the "smart manipulators" usually design the transaction volumes to be "mostly identical" (for example 98%) for standing aside from the regulatory inspections [7]. The difference (Δv) (for example 100%-98%=2%) can then be defined as the *matching margin*. The time intervals between the transactions, shown in the column ***T diff*** in Table 1, can be varied as random events. The transactions occurring at distinct time points may increase the risk of financial loss from price fluctuation, but at the same time, the decrease the possibility of being recognized by the regulator monitoring.

Table 1 AN EXAMPLE OF RING TOPOLOGY WASH TRADE TRANSACTION

Transaction			Seller	Buyer	Volume
ID	Time	T diff (min.)			
T_1	09:00:00	-	A	B	1500 ± Δv
T_2	09:25:00	25	B	C	1500 ± Δv
T_3	09:45:00	20	C	D	1500 ± Δv
T_4	10:20:00	35	D	E	1500 ± Δv
T_5	10:35:00	15	E	F	1500 ± Δv
T_6	10:59:00	24	F	G	1500 ± Δv
T_7	11:25:00	26	G	A	1500 ± Δv

Consider the "parcel passing" trading examples among the collusive traders A, ... , G in Table 1, the transaction T_k between

traders t_i and t_j ($t_i, t_j \in \{A, ..., G\}$) with traded volume $V \pm \Delta v$, can be defined as:

$$T_k = \{t_i, t_j\}_{V \pm \Delta v}, \tag{1}$$

Further, representing the buying and selling actions by positive and negative signs respectively, transaction T_1 (A selling to B $1500 \pm \Delta v$ shares of stock) in Table 1 can be formulated as $T_1 = \{-A, +B\}_{1500 \pm \Delta v}$. The trader with the trading (buying or selling) signs can be called a "signed trader". Similarly, other transactions can be formulated as:

$$T_2 = \{-B, +C\}_{1500 \pm \Delta v},$$

$$\dots$$

$$T_7 = \{-G, +A\}_{1500 \pm \Delta v}.$$

In a wash trade, the transactions among the collusive traders ought to bring no changes in beneficial interest for each party, which means that the overall effect of transactions among the traders in a collusion network equates to nobody buying or selling (no positions change). The effects of two transactions between two collusive traders t_i and t_j add up as follows:

$$\{-t_i, +t_j\}_{V \pm \Delta v} + \{+t_i, -t_j\}_{V \pm \Delta v}$$

$$= \{-t_i + t_i + t_j - t_j\}_{V \pm \Delta v} \tag{2}$$

$$= \{0, 0\}_{V \pm \Delta v} = \{0\}.$$

Based on this, the effect of the transactions in Table 1 can be expressed as

$$\sum_{k=1}^{7} T_k = \sum_{t_i, t_j \in \{A, ..., G\}} \{-t_i, +t_j\}_{V \pm \Delta v} = \{0\} \tag{3}$$

The transactions result in a zero-sum situation where there are no changes in beneficial interest for each party in a ring topologic collusive network.

2) Star topology

The star transaction topology features a "host" trader (A in Fig. 3 left) and a number of "device" traders ($B, ..., G$ in Fig. 3 left) connecting only with host. Since there are no connections among the "device", the "device" traders' positions can be maintained if and only if each "device" trader equalises its sales and purchases with the "host" respectively. The two headed arrows in Fig. 3 illustrate the bi-directional transactions between two neighbours. An example of wash trade following the star topology is shown in Table 2 and is formulated as

$$T_1 = \{-A, +B\}_{1500 \pm \Delta v}, T_2 = \{-B, +A\}_{1500 \pm \Delta v},$$

$$T_3 = \{-A, +C\}_{2500 \pm \Delta v}, T_4 = \{-C, +A\}_{2500 \pm \Delta v},$$

$$\dots$$

$$T_{11} = \{-A, +G\}_{3000 \pm \Delta v}, T_{12} = \{-G, +A\}_{3000 \pm \Delta v}$$

The effect of the transactions can be calculated by summarizing the transactions with the same volume. This involves splitting those transactions into blocks of similar traded volume $V_l \pm \Delta v$ and calculating the summary in each block separately:

$$\sum_{k=1}^{14} T_k = \sum_{t_i, t_j \in \{A, ..., G\}} \{-t_i, +t_j\}_{V_l \pm \Delta v}$$

$$= \{-A, +B\}_{1500 \pm \Delta v} + \{-B, +A\}_{1500 \pm \Delta v}$$

$$+ \{-A, +C\}_{2500 \pm \Delta v} + \{-C, +A\}_{2500 \pm \Delta v} \tag{4}$$

$$\dots$$

$$+ \{-A, +G\}_{3000 \pm \Delta v} + \{-G, +A\}_{3000 \pm \Delta v}$$

$$= \{0\}_{1500 \pm \Delta v} + \{0\}_{2500 \pm \Delta v} + \dots + \{0\}_{3000 \pm \Delta v}$$

$$= \{0\}$$

It is obviously that: if the traders arrange a wash trade following the star network topology, their transactions between any two connected neighbours simply follow a ring topology with two traders (e.g., $A \leftrightarrow B$ as shown in Fig. 3 right part). Therefore, the star topology can be also observed as combinations of a number of "two-trader" ring topologies.
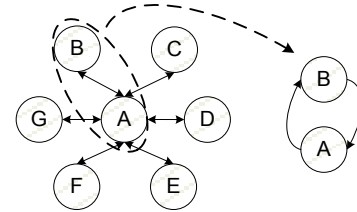

Fig. 3 Star topology with seven traders (left); equivalent "2-trader" ring topology (right).

Table 2 AN EXAMPLE OF STAR TOPOLOGY WASH TRADE TRANSACTION

ID	Transaction		Seller	Buyer	Volume
	Time	T diff (min.)			
T_1	09:00:00	-	A	B	$1500 \pm \Delta v$
T_2	09:25:00	25	B	A	$1500 \pm \Delta v$
T_3	09:45:00	20	A	C	$2500 \pm \Delta v$
T_4	11:25:00	26	C	A	$2500 \pm \Delta v$
	...				
T_{11}	15:25:00	-	A	G	$3000 \pm \Delta v$
T_{12}	16:00:00	35	G	A	$3000 \pm \Delta v$

3) Tree topology

Basic tree topology is based on the hierarchy of nodes as shown in Fig. 4(a). The highest level of any tree consists of a single "root" node (A in Fig. 4(a)) connected to either single or multiple nodes in the level below by point-to-point links. The lower level nodes are also connected to single or multiple nodes in the next level down [9] [10] and the nodes at lowest level are termed as "leaf" nodes (D, E, F and G in Fig. 4(a)).

It is difficult to finish a wash trade based only on the tree topology, because, within a hierarchy structure, the "root" trader can only sell to the traders at lower level while the "leaf" traders can only buy from the higher levels, thus they cannot maintain unchanged trading positions. For example, in the basic tree topology in Fig. 4(a), assuming each transaction has

the same traded volumes, the transactions between the connected nodes can be formulated as

$$T_1 = \{-A, +B\}_{V \pm \Delta v}, \quad T_2 = \{-B, +D\}_{V \pm \Delta v},$$

$$T_3 = \{-B, +E\}_{V \pm \Delta v}, \quad T_4 = \{-A, +C\}_{V \pm \Delta v},$$

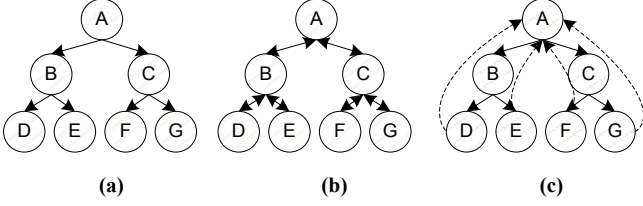$$T_5 = \{-C, +F\}_{V \pm \Delta v}, \quad T_6 = \{-C, +G\}_{V \pm \Delta v}.$$

Fig. 4 (a) Tree topology of seven traders; (b) Tree topology of seven traders with bi-directional paths with; (c) Tree transaction topology of seven traders with loop paths.

According to Equation (3), the total of those transactions is

$$\sum_{k=1}^{6} T_k = \{-2A - B + D + E - C + F + G\}_{V \pm \Delta v} \quad (5)$$

To successfully accomplish the wash trade, more transactions are required to balance the remaining positions. There are two solutions for this. The first one is to simply add six more transactions as:

$$T_7 = \{-B, +A\}_{V \pm \Delta v}, \quad T_8 = \{-D, +B\}_{V \pm \Delta v},$$

$$T_9 = \{-E, +B\}_{V \pm \Delta v}, \quad T_{10} = \{-C, +A\}_{V \pm \Delta v},$$

$$T_{11} = \{-F, +C\}_{V \pm \Delta v}, \quad T_{12} = \{-G, +C\}_{V \pm \Delta v}.$$

Aggregating the six transactions with $\sum_{k=1}^{6} T_k$, the sum is then changed to:

$$\sum_{k=1}^{6+6} T_k = \sum_{k=1}^{6} T_k + \sum_{k=7}^{12} T_k$$

$$= \{-2A - B + D + E - C + F + G\}_{v \pm \Delta v} \quad (6)$$

$$+ \{2A + B - D - E + C - F - G\}_{v \pm \Delta v}$$

$$= \{0\}_{V \pm \Delta v}.$$

If illustrated in Fig. 4(a), the six added transactions actually change the tree connection to bi-directional as shown in Fig. 4(b).

The second solution is to remove the assumption of "same traded volumes" and maintain the balanced positions for the traders at intermediate level (e.g., B and C in Fig. 4(a)). Thus the transactions $T_1, \dots, T_6$ are modified as:

$$T_1 = \{-A, +B\}_{2V \pm \Delta v}, \quad T_2 = \{-B, +D\}_{V \pm \Delta v},$$

$$T_3 = \{-B, +E\}_{V \pm \Delta v}, \quad T_4 = \{-A, +C\}_{2V \pm \Delta v}, \quad (7)$$

$$T_5 = \{-C, +F\}_{V \pm \Delta v}, \quad T_6 = \{-C, +G\}_{V \pm \Delta v}.$$

An aggregate operation for two transactions with at least one identical "signed trader" (same trader with same buy/sell

sign refers to identical trader e.g. $\{+A\}$ is identical to $\{+A\}$ but not to $\{-A\}$) can be written as:

$$\hat{T} = T_{k_1} + T_{k_2}$$

$$= \{+t_i, -t_j\}_{V_1 \pm \Delta v} + \{+t_i, -t_j\}_{V_2 \pm \Delta v} \quad (8)$$

$$= \left\{ [+t_i]_{V_1 + V_2 \pm \Delta v}, [-t_j]_{V_1 + V_2 \pm \Delta v} \right\}.$$

For example, the aggregation $\hat{T}_{2,3}$ of transactions T_2 and T_3 in Equation (7) is shown as follows ($\{-B\}$ is the identical element):

$$\hat{T}_{2,3} = T_2 + T_3$$

$$= \{-B, +D\}_{V \pm \Delta v} + \{-B, +E\}_{V \pm \Delta v} \quad (9)$$

$$= \{[-B]_{2V \pm \Delta v}, [+D, +E]_{V \pm \Delta v}\}.$$

The summary of the initial six transactions with the aggregation operation is then calculated as:

$$\sum_{k=1}^{6} T_k = \{[-A]_{2V \pm \Delta v}, [+B - B]_{2V \pm \Delta v}, [+D, +E]_{V \pm \Delta v}\}$$

$$+ \{[-A]_{2V \pm \Delta v}, [+C - C]_{2V \pm \Delta v}, [+F, +G]_{V \pm \Delta v}\}$$

$$= \{[-A]_{2V \pm \Delta v}, [+D, +E]_{V \pm \Delta v}\}$$

$$+ \{[-A]_{2V \pm \Delta v}, [+F, +G]_{V \pm \Delta v}\}$$

Re-organizing the remaining positions, we can get

$$\{[-A, +D]_{V \pm \Delta v}, [-A, +E]_{V \pm \Delta v}\}$$

$$+ \{[-A, +F]_{V \pm \Delta v}, [-A, +G]_{V \pm \Delta v}\} \quad (10)$$

It is obvious from the result: to balance the remaining position, four more transactions are required as:

$$T_7 = \{+A, -D\}_{V \pm \Delta v}, \quad T_8 = \{+A, -E\}_{V \pm \Delta v},$$

$$T_9 = \{+A, -F\}_{V \pm \Delta v}, \quad T_{10} = \{+A, -G\}_{V \pm \Delta v}. \quad (11)$$

Adding the transactions in Equation (11) to the remaining positions in Equation (10), the summary of the transactions is changed to zero. Again, if illustrating $T_7, \dots, T_{10}$ in Fig. 4(a), the basic tree topology is actually updated with four loop paths (illustrated as the dotted lines in Fig. 4(c)) connecting the "leaf" with the "root".

Consequently, to construct a wash trade, the traders in a tree topological network have to either bi-directionally transact under the tree hierarchy connection or trade with loop paths connecting the "root" and "leaf" nodes together. However, both formats implicitly change the tree topology to a combination of ring topologies with either two traders as Fig. 5(a) or multiple traders as Fig. 5(b).
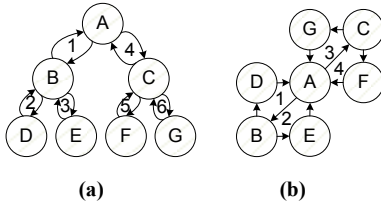
Fig. 5 (a) Tree topology of seven traders with bi-directional paths equals to six "2-trader" ring topologies; (b) Tree topology of seven traders with loop paths equal to four "3-trader" ring topologies.

4) Mesh topology

Mesh topology is a fully connected network, in which each node is connected to each other as shown in Fig. 6. The fully connected network is usually viewed as hybrid network of ring and star topology [9] [10]. Therefore, a wash trade conducted by the traders in a mesh topological network also follows the analysis of ring and start topology in Section III.A.1) and III.A.2).
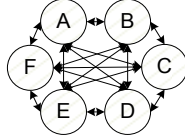

Fig. 6 Mesh transaction topology with seven traders.

As a conclusion, ring topology with one or more traders is the basic structure of the collusive network in wash trade regardless of the complexity of the network, which might be a combination of two or more other topologies. The ring topology can also be observed graphically from the figures: only the transactions with loopback flows (both sale and purchase) are possible to constitute a wash trade. This also conforms to the wash trade condition, no changes in beneficial interest of each party, which is formulated in our analysis as $\sum_{k=1}^{n} T_k = \{0\}$, where n is the number of collusive transactions.

Hence, we choose the ring topology as the indicator of a wash trade action: detecting a certain collection of transactions S so that $\sum_{k \in S} T_k = \{0\}$, triggers a wash trade alert.

B. Wash trade detection

1) Wash trade detection algorithm

Based on the formulation of wash trade topology in Section III.A, the wash trade detection can be formulated as follows: we are given K transactions, $T_k = \{t_i, t_j\}_{V_k \pm \Delta v}, k = 1, \dots, K$ where t_i and t_j represent traders and $V_k \pm \Delta v$ is the traded volume in transaction k (Δv is the *matching margin*). We would like to select a subset S of the transactions so that the summary W of the transactions is zero $W = \sum_{k \in S} T_k = \{0\}$. The number of transactions in the subset S is n ($n \in [1, K]$).

In essence, the formulated problem is a practical case of a more general problem called the *Knapsack Problem* [11] [12] [13]. The name *Knapsack* refers to the problem of filling a knapsack of capacity W as full as possible using a subset of n items $\{1, \dots, n\}$. The wash trade problem can be viewed as a special format of the *Knapsack Problem*: filling a knapsack of capacity W using a subset S of n transactions T_k, each having

negative or positive capacity (e.g. $\{+A\}$ and $\{-A\}$), to make $W=0$, where the calculations of the transaction T_k are defined in equations (3), (4) and (8) and can be easily implemented by operator overloading in C++. The subset S is considered a trading collusion in a wash trade.

Because of the similarity of the two problems, the widely used approach for solving the *Knapsack Problem* [11] [12] [13], dynamic programming, is deployed for the wash trade detection problem. The main principles of dynamic programming is that: we have to come up with a number of sub-problems so that each sub-problem can be solved easily from "smaller" sub-problems, and the solution of the original problem can be obtained easily once we know the solutions to all the sub-problems [14].

To solve the special form of the Knapsack Problem under K transactions and capacity W, denoting the final subset of transactions in an optimum solution for the original problem as S_K, we then use the notation **OPT**(K,W) to denote the summary of the transactions in the subset S as **OPT**$(K,W) = min \left| \sum_{k \in S_K} T_k \right|$. To find out **OPT**$(K,W)$, we not only need the solution of **OPT**$(K\text{-}1,W)$, but also need to know OPT$(K\text{-}1,W - \text{T}_K)$, the best solution on the first K-1 transactions $\{1, \dots, K-1\}$ and the remaining capacity $W\text{-}\text{T}_K$: if T_K is not one of the transactions in the final subset S_K, we can simply ignore the transaction K and find out **OPT**$(K\text{-}1,W)$; but if T_K is, we need to seek an optimal solution in the remaining transactions $1, \dots, K-1$, which is **OPT**$(K\text{-}1,W\text{-}T_K)$. Therefore, the recursion is summarized as

1) If $T_K \notin S_K$, then **OPT**(K,W)=**OPT**$(K\text{-}1,W)$;

2) If $T_K \in S_K$, then **OPT**(K,W)=T_K+**OPT**$(K\text{-}1,W\text{-}T_K)$.

We are therefore going to use many sub-problems: one for each set $\{1, \dots, i\}$ ($i = K, K-1, \dots, 1$) of the transactions and each possible remaining available capacity w. Hence, the summary of the transactions in subset S_i of the optimal solution using transactions $\{1, \dots, i\}$ with the allowed capacity w is denoted as **OPT**$(i,w) = min \left| \sum_{k \in S_i} T_k \right|$ where the capacity w satisfies $\sum_{k \in S_i} T_k = w$. Using this set of sub-problems, we will be able to express the value **OPT**(i,w) as a simple expression in terms of values from smaller problems. This recursive process is re-organized based on the above two conditions to give Algorithm 1, where array T is the given transactions. This recursive algorithm can be used by simply invoking **OPT**(K,W) for K transactions and the capacity W.

Algorithm 1 Wash trade detection algorithm.

1	$T = [T_1, \dots, T_K]$;	// original K transactions;
2	$S = \emptyset$;	// solution subset, initialized to empty;
3	**OPT**(i , w)	// i starts from K and decreases on each recursion step;
4	**if** $i<0$	// if i reaches the last one,
5	**return**;	// then return;
6	**if** $T_i = w$	// if $T_i = w$, the orders in S is one solution;
7	**output** S, T_i;	// output the found solution;
8	**push** T_i into S;	// assume $T_i \in S$,
9	**OPT**$(i\text{-}1,w\text{-}T_i)$;	// recursively find solution by condition 2);
10	**pop** T_i out S;	// assume $T_i \notin S$,
11	**OPT**$(i\text{-}1,w)$;	// recursively find solution by condition 1);
12	**end** of **OPT**;	
13	**return** S;	

2) Pre-organizing: volume bucketing

To apply the algorithm on the detection problem in a practical context, we need to pre-organize the transaction records. As we discussed before, the transactions in wash trade have "nearly identical" ($V \pm \Delta v$) traded volumes, otherwise, some traders in the collusive network would incur a loss or face unnecessary risk, which is not accepted by the manipulators. Therefore, we argue that for detecting the wash trade, the traded volume is a more relevant metric to sort the transaction records than the physical time, although the transactions are usually recorded as a time sequence. In essence, the sorting involves splitting trading sessions into blocks of traded volume for the purpose of detecting the potential collusions in each interval.

However, the original trading records are not appropriate for the sorting due to large masses of trading noises from the active speculators issuing large number of small orders in one single day, which makes the transaction sequence in a mess. The wash trade is to manipulate the trading volume, the orders with tiny sizes from the noise traders are surely not within the detection scope. According to the empirical studies in [7], the traders involved in the volume manipulations (e.g. wash trade) of one certain stock usually do not trade legitimately within a certain time period which is usually one trading day. That means if trader "A" and "B" are manipulators of a wash trade, there is a little chance that they also trade normally on the same stock within one trading day. Therefore, we aggregate the transactions with at least one identical "signed trader" as defined in equation (8) and then set up the volume buckets for sorting the transactions into different volume buckets. The detection process in Algorithm 1 will be applied within each bucket separately.

Table 3 WASH TRADE TRANSACTION EXAMPLE IN TREE TOPOLOGY WITH LOOP PATHS

Transaction ID	Time	Seller	Buyer	Volume
1	09:00:00	A	B	$2000 \pm \Delta v$
2	09:25:00	B	D	$1000 \pm \Delta v$
3	09:45:00	B	E	$1000 \pm \Delta v$
4	11:25:00	D	A	$1000 \pm \Delta v$
5	15:25:00	E	A	$1000 \pm \Delta v$
6	16:00:00	A	C	$2000 \pm \Delta v$
7	16:00:00	C	F	$1000 \pm \Delta v$
8	16:00:00	C	G	$1000 \pm \Delta v$
9	16:00:00	F	A	$1000 \pm \Delta v$
10	16:00:00	G	A	$1000 \pm \Delta v$

Table 3 illustrates the transaction examples in a tree topology with loop paths as shown in Fig. 4(c). Based on the calculation in equation (8), the trading records can be aggregated as:

$$\hat{T}_{1,6} = T_1 + T_6 = \{[-A]_{4000}, [+B, +C]_{2000}\},$$

$$\hat{T}_{2,3} = T_2 + T_3 = \{[-B]_{2000}, [+D, +E]_{1000}\},$$

$$\hat{T}_{4,5} = T_4 + T_5 = \{[-D, -E]_{1000}, [+A]_{2000}\},$$

$$\hat{T}_{7,8} = T_7 + T_8 = \{[-C]_{2000}, [+F, +G]_{1000}\},$$

$$\hat{T}_{9,10} = T_9 + T_{10} = \{[-F, -G]_{1000}, [+A]_{2000}\}.$$

The buckets are selected as $[1000, 4000]$ based on the aggregated traded volumes. By this, the trading records (orders) are reorganized and pre-filtered by the aggregated trading volumes and then Algorithm 1 will be applied on the volume buckets for detecting the potential wash trade collusion.

The analogous volume buckets approach is also used in financial market volatility analysis [15].

IV. EXPERIMENTS AND EVALUATION

Evaluating a detection algorithm usually relies on real data of both "normal" and "abuse" cases. However, due to the limited reports of wash trade manipulation and regulatory rules prohibiting the disclosure of illegitimate market data, the availability of the examples of wash trade in capital market is more limited than that of routine normal trading records. Due to the absence of real market wash trade cases, it is acceptable to the financial industry business that all the characteristic patterns of wash trade examples are studied and then reproduced in other financial data context to synthesize artificial manipulation cases [16]. Synthetic exploratory financial data are also accepted in academia for evaluating the proposed algorithm when real market data are hard to collect [5] [17] [18]. Thus in this approach, the wash trade scenarios are synthetically generated following the analysis in Section III and injected in other real normal trading transaction streams.

The experimental data used in this work involves real market data of four stocks including Google (GOOG), Microsoft (MSFT), Intel (INTC) and Apple (AAPL) from NASDAQ. These datasets are selected due to their active trading activities, relatively high trading volumes and more volatile price fluctuation, the factors that might increase the likelihood of market abuse across the exchanges [19] [1]. The datasets, obtained from LOBSTER project [20], covers transactions over five trading days 11th -15th June 2012 and consists of more than 400,000 records.

The typical wash trade scenarios are reproduced within each stock context. The scenarios are reproduced according to the five topological structures: single trader (as Fig. 1), ring (as Fig. 2), star (as Fig. 3) and tree (with bi-directional and loop path as Fig. 4 middle and right). One hundred wash trade examples are generated for each topology respectively and each example comprises of a number of transactions. Thus 5*100=500 wash trade examples are synthetically generated and injected to the normal transaction records of the corresponding stocks making the test data a mixture of both "normal" and "wash trade" patterns.

The performance evaluation of the proposed algorithm is based on two popular statistical measures: sensitivity (*SEN*) and specificity (*SPE*). Both of them are based on the confusion matrix, where false positive, *FP*, is defined as wash trade cases detected as normal and false negative, *FN*, is defined as normal cases detected as a wash trade, true positive, *TP*, is defined as normal cases detected as normal and true negative, *TN*, is defined as wash trade cases detected as wash trade. The sensitivity, defined as *SEN* = *TP* /(*TP* + *FN*), represents the rate of correctly detecting normal trading orders while the specificity, defined as *SPE* = *TN* /(*TN*+*FP*), refers to the rate

of precisely figuring wash trade cases. The sensitivity and specificity measure across four stocks with injected wash trade examples are illustrated in Table 4.

Table 4 SENSITIVITY AND SPECIFICITY OF FOUR STOCKS FROM NASDAQ WITH INJECTED WASH TRADE CASES

stock	specificity	sensitivity
GOOG	100.00%	99.96%
MSFT	100.00%	99.86%
INTC	100.00%	99.20%
AAPL	100.00%	99.94%

The specificity values show that in the datasets of four stocks, the injected wash trade collusive transactions are all successfully detected by the proposed algorithm. However, the sensitivity values vary across different stocks. The stock INTC achieves the lowest value as 99.20%. After carefully checking the false negative cases (normal cases detected as a wash trade) in INTC dataset, we found that there are a number of transactions between certain traders across different days as two examples shown in Table 5.

Table 5 FALSE NEGATIVE TRANSACTIONS OF STOCK INTC

case #	time (US Easter)	volume	seller ID	buyer ID
1	11/06/2012 16:29:55	13559	CLIENT12	CLIENT3
	12/06/2012 10:53:54	13356	CLIENT3	CLIENT12
2	14/06/2012 16:00:51	18328	CLIENT8	COLLINSS
	15/06/2012 09:29:08	18300	COLLINSS	CLIENT8

In case #1, the trader CLIENT12 sold 13559 shares to CLIENT3 at market closing time and bought 13356 shares back on the next morning. The traders CLIENT8 and COLLINSS traded similarly with nearly identical volumes (18328 and 18300 shares). Although not reported by regulators, the two cases are very analogous to wash trade based on the existing regulatory rules from FSA and CESR, which though do not indicate the timing features (inter-day or intra-day) of wash trade. Given the facts that the reported wash trade cases usually occurred as an intra-day trading and there is also no theoretical study and regulatory rules on the trading time of the wash trade, it is extremely hard to determine the legality of the false negative cases by the proposed algorithm, which rather aims at triggering alerts about suspicious collusive transactions.

V. CONCLUSION

A wash trade activity detection algorithm is proposed in this paper. To derive the algorithm, the collusive activities in the wash trade are analysed by using four network topologies: ring, star, tree and mesh. The analysis demonstrates the ring topology as the basic structure of the transactions in a wash trade. Based on the analysis, a recursive algorithm is proposed addressing the detection problem. The algorithm is applied on the aggregated and volume-sorted transaction sequences. The experimental evaluations, conducted on the trading records of four real stocks with injected synthetic wash trade transactions, show the effectiveness of the algorithm.

Some false negative cases showed the wash trade features but occurred across two trading days. From the detection model's perspective, those false negative cases can be easily bypassed by setting up the detection time period to be one single trading day. However, this requires more empirical or theoretical studies on the time characteristics of wash trade, which is a major issue that requires more future investigations.

In recent years, the wash trade manipulation tends to be carried out in more than one exchange market by some tricky manipulators. Detection within any single market hardly achieves a complete and accurate result. This requires a cross-market detection model, which is one of our primary future works.

REFERENCE

[1] D. J. Cumming, F. Zhan and M. J. Aitken, "High Frequency Trading and End-of-Day Manipulation," Social Science Research Network, 2012.

[2] FSA, "The Code of Market Conduct," March 2006. [Online]. Available: http://www.fsa.gov.uk/pubs/hb-releases/rel52/rel52mar.pdf.

[3] CESR, "Market Abuse Directive," THE COMMITTEE OF EUROPEAN SECURITIES REGULATORS, Paris, 2005.

[4] M. Franke, B. Hoser and J. Schröder, "On the Analysis of Irregular Stock Market Trading Behavior," in *Data Analysis, Machine Learning and Applications*, Springer-Verlag Berlin Heidelberg, 2008, pp. 355-362.

[5] G. K. Palshikar and M. M. Apte, "Collusion set detection using graph clustering," *Data Mining and Knowledge Discovery,* vol. 16, no. 2, pp. 135-164, 2008.

[6] J. Wang, S. Zhou and J. Guan, "Detecting potential collusive cliques in futures markets based on trading behaviors from real data," *Neurocomputing,* vol. 92, pp. 44 - 53, 2012.

[7] M. Aitken, F. R. Harris and S. Ji, "Trade-based manipulation and market efficiency: a cross-market comparison," in *22nd Australasian Finance and Banking Conference*, 2009.

[8] NANEX, "Chicago PMI," 31 May 2013. [Online]. Available: http://www.nanex.net/aqck2/4304.html.

[9] R. Brown, Topology and Groupoids, BookSurge Publishing, 2006.

[10] D. Groth and T. Skandier, Network+ Study Guide, San Francisco, London: SYBEX, 2005.

[11] R. Andonov, V. Poirriez and S. Rajopadhye, "Unbounded knapsack problem: Dynamic programming revisited," *European Journal of Operational Research,* vol. 123, no. 2, p. 394–407, 2000.

[12] V. Poirriez, N. Yanev and R. Andonov, "A hybrid algorithm for the unbounded knapsack problem," *Discrete Optimization,* vol. 6, no. 1, p. 110–124, 2009.

[13] M. Zukerman, L. Jia, T. Neame and G. J. Woeginger, "A polynomially solvable special case of the unbounded knapsack problem," *Operations Research Letters,* vol. 29, no. 1, pp. 13-16, 2001.

[14] J. Kleinberg and E. Tardos, Algorithm Design, Addison-Wesley, 2005.

[15] D. Easley, M. d. Prado and M. O'Hara, "Flow toxicity and volatility in a high frequency world," Johnson School Research Paper Series 9, 2011.

[16] NANEX, "Exploratory Trading in the eMini," NANEX, 10 July 2013. [Online]. Available: http://www.nanex.net/aqck2/4136.html.

[17] T. B. Ho and Z. H. Zhou, "Domain-Driven Local Exceptional Pattern Mining for Detecting Stock Price Manipulation," in *PRICAI 2008: Trends in Artificial Intelligence*, Hanoi, 2008.

[18] M. Franke, B. Hoser and J. Schröder, "On the Analysis of Irregular Stock Market Trading Behavior," in *Data Analysis, Machine Learning and Applications*, Freiburg, 2007.

[19] E. J. Lee, K. S. Eom and K. S. Park, "Microstructure-based manipulation: Strategic behavior and performance of spoofing traders," *Journal of Financial Markets,* vol. 16, no. 2, pp. 227 - 252, 2013.

[20] LOBSTER, "LOBSTER," Humboldt Universität zu Berlin, 2013. [Online]. Available: https://lobster.wiwi.hu-berlin.de/index.php.