Detecting Motifs in Multiplex Corporate Networks

Frank W. Takes, Walter A. Kosters and Boyd Witte

Abstract The main topic of this paper is the discovery of motifs in multiplex corporate networks. *Network motifs* are small subgraphs occurring at significantly higher numbers than in similar random networks. They can be seen as the building blocks of a complex network. In real-world network data, multiple types of (possibly overlapping) relationships may be present among the nodes, forming so-called *multiplex networks*. Detecting motifs in such networks is difficult, as existing subgraph enumeration algorithms are not directly applicable to multiplex network data. In addition, the selection of a proper multiplex null model to test the significance of the enumerated subgraphs is nontrivial. This paper addresses these two problems, resulting in three contributions.
First, we present a method based on layer encoding for adequately handling the multiplex aspect in subgraph enumeration. Second, a null model is proposed that is able to preserve the relationship between the different types of links, taking into account that a particular link type may be the result of a projection from a bipartite network. Finally, we perform experiments on *corporate network* data from Germany, in which around 75 000 nodes represent corporations and roughly 195 000 links represent connectedness of firms based on shared board members and ownership. We demonstrate how incorporating the multiplex aspect in motif detection is able to reveal new insights that could not be obtained by studying only one type of relationship. Furthermore, results uncover how the financial sector is over-represented in the more complex motifs, hinting at a surprisingly prominent role of the financial sector in the largely industry-oriented corporate network of Germany.

Frank W. Takes
CORPNET Research Group (AISSR), University of Amsterdam, e-mail: `takes@uva.nl`
Leiden Institute of Advanced Computer Science (LIACS), Leiden University

Walter A. Kosters
Leiden Institute of Advanced Computer Science (LIACS), Leiden University,
e-mail: `w.a.kosters@liacs.leidenuniv.nl`

Boyd Witte
Leiden Institute of Advanced Computer Science (LIACS), Leiden University

1 Introduction

The structure of networks, in which nodes represent objects that are connected through particular relationships, is frequently analyzed [3]. Typically, network analysis deals with trying to understand how links at the micro (object and its relationships) level result in a particular system behavior at the macro (network) level. For example, in social systems, friendships between people at the micro level result in a network of friendships exhibiting small world properties such as low average distances at the macro level. In this paper we are concerned with the intermediate level, the so-called meso level of networks. At this level, we consider network motifs; small building blocks of a handful of nodes and edges that are characteristic for the particular system that is considered [2, 20].

The main network data considered in this paper is that of corporate networks. In these networks, nodes represent corporations and links between nodes indicate that the two firms have a particular economic relationship. The key element here is that there is usually more than one type of relationship that binds two firms. In our case, we have data on ownership relations, where a firm owns a significant portion of another firm, allowing this firm to exert control over the other. In addition, we know the compositions of the boards of directors, i.e., we know who serves as a senior level director at which firm. As many directors hold more than one position, so-called board interlock links are formed, connecting firms if they share one or more senior level directors. Both types of links have been shown to be extremely relevant. Ownership networks reveal for example which firms have control in our global economy [25] and how network structures via particular countries are used for tax evasion [10]. Board interlock networks have important consequences for corporate governance [19] and have been instrumental in revealing the community structure of a business elite governing larger corporations across the globe [14].

The observation that nodes in a complex network are connected through multiple types of links has resulted in ample studies on so-called multiplex networks, sometimes also called multi-layer networks [15]. The aim of these studies is often to take advantage of the multiple types of relationships that exist, obtaining insights that could not be discovered if only one type of relationship was considered. Here, we aim to do this in light of the discovery of network motifs. In our corporate network data, multiplex motifs may shed new light on the organization of businesses in networks, and how firms exert power and control over other firms.

However, enumeration of motifs in a multiplex network provides methodological challenges, as state-of-the-art methods for efficient subgraph enumeration are not directly applicable to multiplex networks [21, 22, 26]. Furthermore, in order to test if a particular pattern is significantly more frequently occurring and thus not random, the results need to be quantitatively compared against a null model. Another challenge in multiplex motif detection is that there is typically interlayer assortativity [8], meaning that layers of the network are not randomly connected, but according to some pattern of assortativity in which nodes with similar properties (e.g., degree) are more likely to connect to each other in multiple layers. Our research is furthermore complicated by the fact that the board interlock relationships actually result from a projection from underlying bipartite network data, which further complicates the combination of different layers of the network into a null model.

This paper addresses the aforementioned difficulties by providing three contributions. First, we describe a novel adaptation of an existing subgraph enumeration algorithm based on layer encoding, which allows the multiplex subgraphs to be enumerated. The second contribution is a suitable null model that is able to handle both interlayer assortativity and the fact that networks are projected variants of bipartite networks. Third, using data on the largest firms in Germany, we demonstrate how multiplex motif detection is able to reveal new insights, in particular with regards to the involvement of particular economic sectors in more complex motifs.

The remainder of this paper is organized as follows. First, in Section 2 we discuss relevant previous and related work. Next, Section 3 introduces the necessary terminology to formulate the formal problem statement. Then in Section 4 we describe our network dataset and its layers, after which the main approach of multiplex motif detection is outlined in Section 5. In Section 6 we test the approach on our data, resulting in a number of conclusions and suggestions for future work in Section 7.

2 Previous work

Here we discuss related work on motifs, corporate networks and multiplex networks.

Motif recognition has been applied in a number of domains, such as biological networks [18], brain networks [16, 4] and social networks [6]. It is interesting from a computational perspective, because enumerating subgraphs is a #P-hard task [21], as each subset of nodes in a graph has to be compared against all known (possibly isomorphic) subgraphs. Therefore any motif recognition algorithm must be provided with a small enough input, or give an approximation [21]. Some algorithms address this complexity problem by requiring the user to provide input on which subgraphs should be counted or what threshold the frequency of the motif should pass [12]. Other motif recognition algorithms work around the computational limitations by only finding a specific subset of patterns, such as dense subgraphs [13]. Methods like G-TRIES [21], FANMOD [26], and SUBENUM [22] find only induced subgraphs, the type of subgraphs that we also consider in this paper. Apart from motif enumeration, research has also been done on motif counting [6], in which the goal is to count all motifs of a particular type without iterating over all subgraphs consisting of that motif. Although counting algorithms are typically faster, ultimately we are interested in what insights the discovered motifs provide, meaning we need to know their composition. Therefore, our focus is on motif enumeration.

Corporate networks have extensively been analyzed in terms of network topology [25], centrality [24] and community detection [14]. Yet, little to none of this work deals with detecting motifs. In [20], interfirm relationships of materials and services exchanged are analyzed up to patterns of size three, counting V-shaped and triangle-shaped structures. However, to the best of our knowledge there are no studies of (larger) motifs in corporate networks based on ownership and/or interlocks.

Multiplex networks such as our corporate networks in which multiple types of interaction take place, have extensively been studied and classified, see [15] for an overview. Important to note is that we focus on networks in which the same set of nodes that may be connected by different and possibly overlapping types of relationships. These networks are sometimes also called multi-relational, multi-dimensional or multi-layer networks. A good overview of these naming conventions and definitions can be found in [7]. Importantly, the goal of multiplex network analysis is to not lose information by aggregating the layers of the network, taking advantage of the insights that result from the multiple types of interaction [8]. In this light, a number of multiplex network characteristics and methods at the micro and macro level have been devised, including community detection and centrality [23]. This work contributes to the field of multiplex network analysis by means of a method of analysis at the meso level: multiplex motif recognition.

3 Terminology and problem statement

This section introduces formal definitions of network motifs and multiplex networks, leading to the precise specification of multiplex motif detection.

3.1 Networks and motifs

A *graph* or *network* $G = (V,E)$, consists of a set of nodes $V = V(G)$ (also called objects or vertices) and a set of directed edges $E = E(G) \subseteq V \times V$ (also called relationships or links). Nodes are identified using some unique identifier (ID) or label. We assume that there are no parallel edges or self-loops. A graph g is a *subgraph* of graph G if and only if $E(g) \subseteq E(G)$ and $V(g) \subseteq V(G)$, where all nodes incident with an edge in $E(g)$ occur in $V(g)$. A subgraph g is an *induced subgraph* of G if for any pair of nodes $u, v \in V(g)$, it holds that if $(u,v) \in E(G)$ then $(u,v) \in E(g)$. We only consider *connected induced subgraphs* in which all nodes are (indirectly) linked through edges, ignoring link direction. The size k of a subgraph g is its node count, i.e., $k = |V(g)|$.

The *pattern* of a (sub)graph is its abstract representation without particular identifiers or labels. All isomorphic (sub)graphs thus have the same pattern. Let I denote the collection of all patterns. We define $S^i(G)$ as the set of subgraphs of pattern $i \in I$ in graph G. The *frequency* of pattern $i \in I$, denoted $|S^i(G)|$, is the number of occurrences of pattern i in graph G. A *motif* is a pattern that is considered significant according to a particular frequency-based comparison or metric (as further discussed in Section 5.2). The set of all motifs of size k in graph G is denoted $M_k(G)$, and $M(G) = \cup_k M_k(G)$.

3.2 Multiplex networks and motifs

A *multiplex graph* (or *network*), denoted $\mathscr{G} = (V,E,J)$, is a graph that contains multiple types of edges. The collection of edge types is called J. We use $E_j(\mathscr{G})$ with $j \in J$ to refer to the set of edges of type j. There is at most one edge of a certain type in the same direction between any two nodes, meaning that if there are multiple edges between two nodes, they are of different types. In a *multiplex induced subgraph* g it holds that for any pair of nodes $u,v \in V(g)$ in subgraph g and for each type of edge $j \in J$ that if $(u,v) \in E_j(\mathscr{G})$ then $(u,v) \in E_j(g)$. Following similar definitions for patterns and frequency as in Section 3.1 (e.g., introducing $S^i(\mathscr{G})$), a *multiplex motif* is a multiplex pattern that is considered significant according to a particular frequency-based comparison (as further discussed in Section 5.2). The set of all motifs of size k in multiplex graph $\mathscr{G}$ is denoted $\mathscr{M}_k(\mathscr{G})$, and $\mathscr{M}(\mathscr{G}) = \cup_k \mathscr{M}_k(\mathscr{G})$.

3.3 Problem statement

The goal of this paper is, given as input a multiplex graph $\mathscr{G}$ and motif size k, computing the set of multiplex motifs $\mathscr{M}_k(\mathscr{G})$. This problem consists of two tasks: enumerating multiplex subgraphs (as discussed in Section 5.1) and motif significance testing (as elaborated on in Section 5.2). Important to note is that we are not only interested in counting the frequency of motifs, but in enumerating them. The advantage of enumeration is that that we can observe which nodes in the empirical network are involved in the motifs, actually allowing the patterns to be interpreted.

4 Data

The data used in this paper originates from Bureau van Dijk's Orbis database (`http://orbis.bvdinfo.com`). Orbis is a frequently used corporate database, compiled from official country registrars and other collection agencies, often identified as one of the most reliable and complete sources of corporate data [11, 14, 25]. In November 2015 we extracted all active German companies for which ownership and/or board information was available. We also extracted each firm's economic sector. In addition, we extracted all significant ownership relations between these firms with a share of at least 5%, a common threshold at which a stake is considered significant. These relations together form a directed network G_a in which a link $(u,v) \in E_a$ indicates that firm u owns a part of firm v and is thus able to exert control over it. We also extracted for all firms their top executives (chiefs and directors), forming a bipartite network in which directors are connected to firms if the director serves at the board of that firm. This bipartite network can be projected onto an undirected one-mode network G_b in which links $\{u,v\} \in E_b$ indicate that u

Table 1: Network statistics

Network	**Nodes**	**Edges**	**Density**	**Clustering**
Ownership	37 724	31 506	$2.25 \cdot 10^{-5}$	0.033
Board interlock	61 209	175 108	$4.67 \cdot 10^{-5}$	0.384
Multiplex	75 224	195 073	$1.72 \cdot 10^{-5}$	0.277

Table 2: Division of firms over economic sectors

Sector	**Ownership**		**Board interlock**		**Multiplex**	
Bank	474	1.25%	865	1.41%	972	1.29%
Financial	4 648	12.32%	6 250	10.21%	8 338	11.08%
Foundation/Research	55	0.14%	51	0.08%	88	0.12%
Industrial	32 350	85.75%	53 767	87.84%	65 484	87.05%
Insurance	19	0.05%	26	0.04%	34	0.05%
Mutual & Pension Fund	112	0.30%	175	0.29%	213	0.28%
Private Equity	29	0.08%	30	0.05%	37	0.05%
Public Authority	22	0.06%	31	0.05%	41	0.05%
Venture Capital	15	0.04%	14	0.02%	17	0.02%

and v share at least one director. We now have a multiplex network $\mathscr{G}$ with a layer of directed ownership links E_a and a layer of undirected board interlock links E_b.

Table 1 reports basic network statistics such as the number of nodes, links (which are symmetric for the board interlock network), density, and average local clustering coefficient of our corporate network dataset. See [3] for definitions of these metrics. There is significant overlap in the different layers of the multiplex network; $23,709$ nodes (32% of the nodes in the multiplex network) is involved in both ownership and board interlock links. Similarly, $11,541$ edges are what we call multiplex edges; connecting nodes with both an ownership link and a board interlock, constituting 37% of the ownership links, 6.7% of the board interlocks and in total 5.9% of the multiplex network. Finally, the division of firms over economic sectors is shown in Table 2, demonstrating the prominent role of German industrial sector.

5 Approach

First, Section 5.1 explains how an existing state-of-the-art subgraph enumeration algorithm can be adjusted to handle multiplex network data. Next, in Section 5.2 we outline a null model that is suitable for multiple link types. Section 5.3 discusses ways of comparing subgraphs found in the empirical data to the null model.

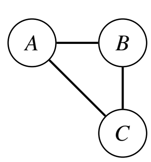
5.1 Multiplex subgraph enumeration

We first briefly discuss the SubEnum [22] algorithm for subgraph enumeration on which our approach is based, and then make the step to multiplex networks.

SubEnum The input of SubEnum is a directed network. The output is a set of subgraphs and their frequencies. Undirected edges are represented as symmetric directed links. At the basis is the Enumerate Subgraph algorithm (ESU) [26], which counts subgraphs in directed unweighted graphs. It loops over all nodes starting at the node with the lowest ID, recursively expanding on every neighboring node with a higher ID until the set of nodes is of size k. The resulting set of nodes including the edges that exist between these nodes, is an induced subgraph, which is then given a canonical label with the Nauty [17] algorithm. This label is guaranteed to be equal for all isomorphic subgraphs. See Figure 1 for a simple example of this labeling step. SubEnum [22] is a parallel variant of the ESU algorithm, solving thread load balancing issues by performing the aforementioned expanding process on edges instead of nodes. To work round memory limitations, it adjusts the way subgraphs are checked for isomorphismm, using a two phase isomorphism check where intermediate results are stored to disk. For a more detailed description, see [22].

Multiplex SubEnum The proposed multiplex adaptation is a two step process: adjusting the subgraph recognition algorithm SubEnum and adjusting the isomorphism detector Nauty accordingly. Here we exploit the fact that any multiplex graph $\mathscr{G}$ can be expressed as a directed labeled graph G'. Instead of explicitly storing multiple edge types, the multiplex graph is converted into a directed labeled graph in which each edge has a label based on the edge types present between the two nodes that it connects. The label consists of a binary string of length J (the number of layers), of which the bit at index i is equal to 1 an edge of type i is present and 0 otherwise (note an ordering is applied to J so an index can be assigned to each edge type). This binary label can be seen as an edge weight, as illustrated in Figure 2a and Figure 2b. It should be noted that this conversion to a seemingly weighted graph G' does not imply that we are suddenly dealing with a weighted graph; we are merely encoding layer presence or absence, and summarize this with a number.

Although we now have weights/labels representing the layers, the original ESU algorithm does not handle labels nor weights. Therefore we propose that when the algorithm encounters a subgraph, a label is created based on the adjacency matrix

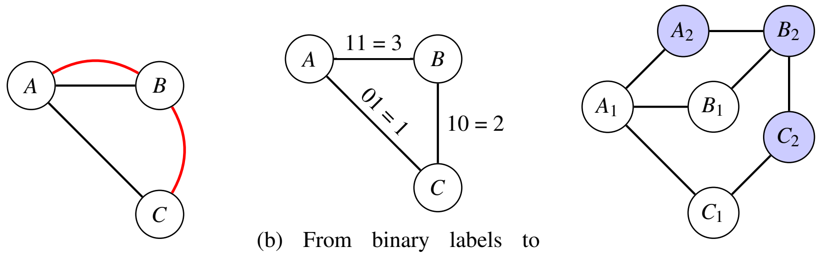


(a) Undirected graph

$$\begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \end{array} \rightarrow 011101110$$

(b) Label of Figure 1a

Fig. 1: Pattern labeling in undirected networks.

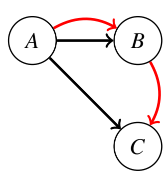


(a) Multiplex edges (b) From binary labels to weighted edges (c) Colored nodes

Fig. 2: From multiplex edges to binary labeled / weighted edges to colored nodes.

with weights representing the layer encoding, as shown in Figure 3. Then, to adapt SUBENUM to handle weighted graphs, we only have to adapt the label constructor so that it incorporates the edge weights.

The second step is adjusting NAUTY to handle the weighted graphs, for which we use node-colored graphs, which have multiple node types (colors). This method is similar to the suggestion for expressing weighted graphs given in NAUTY's documentation [17]. We create a new node-colored graph graph G'' from G', which is the graph with binary labels representing multiplex graph $\mathscr{G}$ as discussed above. The number of node colors is equal to $|J|$, and each color is used to express a single edge type, according to the binary label. For each node in $V(G')$, a set of $|J|$ colored nodes is created in $V(G'')$. So for every node $A \in V(G')$, a set $\{A_1, A_2, \ldots, A_{|J|}\}$ with different colors is added to $V(G'')$. Then, for $1 \leq j < |J|$, every $A_j \in V(G'')$ is connected to A_{j+1} by adding an undirected edge (A_j, A_{j+1}) to $E(G'')$. This creates a string of colored nodes for each node in the original network. Then, crucially, an edge between two nodes A_j and B_j is used to express the presence of the j^{th} edge type encoded in the binary label. An example with undirected edges can be seen in Figure 2c, where the multiplex graph from Figure 2a with two types of edges is shown rewritten with two types of colored nodes.



(a) Directed multiplex graph

$$\begin{array}{cc} & \begin{array}{ccc} A & B & C \end{array} \\ \begin{array}{c} A \\ B \\ C \end{array} & \begin{pmatrix} 0 & 3 & 1 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix} \end{array} \rightarrow 031002000$$

(b) Label of Figure 3a

Fig. 3: Pattern labeling in directed multiplex networks.

5.2 Null model

A suitable null model for multiplex networks must deal with the dependencies between the different layers. For example, in our empirical data, 5.9% of all edges overlap, e.g., there is both an ownership link and a board interlock. A quick calculation shows that as a result of the low density of the networks of each type ($2.25 \cdot 10^{-5}$ and $4.67 \cdot 10^{-5}$, respectively), merging two separately generated random networks will have few overlapping edges. Indeed, the concept of interlayer assortativity [8] is common across different multiplex networks and has to be preserved in the null model. Therefore, a model which fixes degree sequences of both edge types is required. Here we build on the stub-matching model [5], which generates networks with a particular fixed (in and out)degree sequence. To ensure that degree sequences are fixed for all edge types, each combination of edge types is modeled separately, fixing the node degrees for each (combination of) edge type(s). Thus, we model in total $2^{|J|} - 1$ different networks. This is a mere three network models in our case, namely for the ownership links, board interlocks and combined "multiplex edge".

A second challenge is that the board interlock network is the result of a projection of the bipartite network linking firms and directors to a firm-by-firm network, linked based on shared directors. As such, many cliques exist in the empirical network, resulting from directors with three or more positions. To ensure that this property is preserved, the undirected interlock network is modeled at the bipartite level. For this, we again employ the stub-matching model [5]. We encode the node type (firm or director) by enforcing that directors only have a particular outdegree value, and firms only a particular indegree value. The subsequent conversion to an undirected network is trivial, after which a regular projection to the one-mode firm-by-firm network can be made. It should be noted that in our case, the same bipartite projection step should be done for multiplex edges, because part of a multiplex edge is an interlock edge. Finally, the application of this multiplex model allows us to generate a set Y of networks to which the empirical network data can be compared.

5.3 Evaluation metrics

The significance of a pattern (counted subgraph) and classification as a motif is determined using two metrics. These two metrics ensure that we determine pattern significance based on both comparison with the null model, as well as based on the empirical graph alone among patterns of a similar size.

- The *ratio* $R(i,\mathscr{G})$ of a pattern i in graph $\mathscr{G}$ given a set of sampled random multiplex graphs Y (the null model) is defined as follows:

$$R(i,\mathscr{G}) = |S^i(\mathscr{G})| \cdot \left(\frac{\sum_{\mathscr{H} \in Y} |S^i(\mathscr{H})|}{|Y|} \right)^{-1}$$

Table 3: Patterns and motifs per network

	Pattern size				**Motif size**			
	3	**4**	**5**	**All**	**3**	**4**	**5**	**All**
Ownership	11	63	391	465	3	4	6	13
Board interlock	2	6	21	29	0	2	10	12
Multiplex	58	1 132	21 858	23 048	14	48	73	135

When the ratio is larger than 1, the probability of pattern i appearing in the empirical network is larger than the probability of i appearing in a random graph [26].

- The *concentration* $C(i,\mathscr{G})$ of a pattern i (of size $|i|$) in graph $\mathscr{G}$ is the ratio between its frequency and frequencies of all patterns of the same size (see [26]):

$$C(i,\mathscr{G}) = \frac{|S^i(\mathscr{G})|}{\sum_{j\in I,|j|=|i|} |S^j(\mathscr{G})|}$$

6 Experiments

The multiplex motif detection approach explained in Section 5 will be applied to the corporate network dataset in Section 4. The null model that serves as a baseline for the significance of obtained results (see Section 5.2) is generated using $1,000$ samples, as suggested in [26]. As for the evaluation metrics proposed in Section 5.3, we manually set a cut-off value of 5 for the ratio and 0.01 for concentration. Figure 4a shows how these cut-off values capture only the interesting motifs (note the asymmetric logarithmic axes of the figure). To keep computation time within reasonable limits, we run the algorithm up to motif size $k = 5$. An implementation of the approach as well as an exhaustive list of the motifs can be found at the supplementary material website `http://liacs.leidenuniv.nl/~takesfw/multiplexmotifs`.

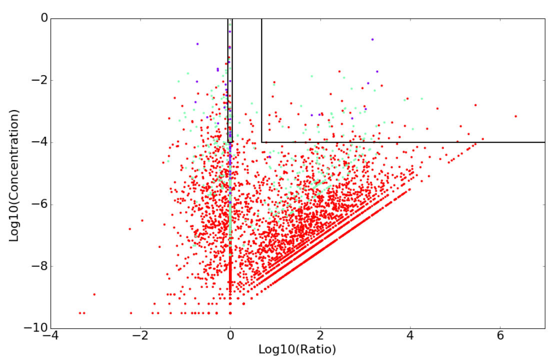
Table 3 shows for increasing motif sizes the number of patterns (enumerated subgraphs) and motifs (significant patterns). The cut-off values of concentration and ratio reduce the 23 048 patterns to only 135 motifs. From these motifs, we highlight a few with high concentration, ratio and interesting economic sector composition.

Size 3. At the ownership level, several of the discovered size 3 motifs reveal the existence of crossholdings, i.e., mutual investment in firms to strengthen a country's internal economy, a common phenomenon in Germany [1]. Furthermore, the motif of size 3 in Figure 5a shows how it is common for investment (ownership links) and control (board interlocks) to go hand in hand. This motif with a ratio of 1032 and concentration of 0.118 shows how investments could be backed by shared directors to exercise control, or how shared directors may have facilitated the investment (recall we do not have temporal information to assess causality).

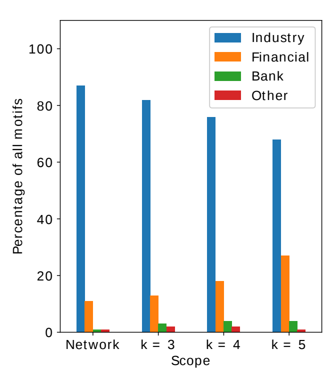
Size 4. Interesting to note is the size 4 motif in Figure 5b, with a ratio of 2024 and concentration of 0.351. The division over economic sectors in Table 2 shows that 87% of the firms is in the industrial sector. Yet, this motif's links are between 43% industrial and 56% financial firms. It reflects a common corporate structure of two joint ventures by the same investors. The size 4 motif with ratio 9 and concentration 0.717 in Figure 5c shows a board interlock and two separate investments. Board interlocks have been intensively studied, and an important motivation for firms to establish an interlocks is monitoring. Originally, this was largely done by banks for reasons of oversight [9, 19], ensuring that their diverse investments were properly managed. This is reflected by the motif, in which 30% of the links involves a bank, whereas banks only comprise 1.29% of all firms.

Size 5. Figure 5d shows a motif of size 5 with one of the highest ratio values, namely 113 400. Interestingly, it turns out that "Mutual & Pension Fund" firms are often involved in this motif. Whereas only 0.28% of the firms in the data is of this type, 14% of edges in this particular motif involve such a firm. The structure represents two investments into two firms governed by the same director. Indeed, it is common for funds not to randomly invest, but to strategically choose firms at which one knows a particular board member from a previous investment. Interesting to note is that the size 3 version of this motif, the pattern of investment by one firm in two firms with shared directors, does not have such an over-representation of these types of firms. This confirms that the unique aspect is the multiple investments, and in general shows how looking at larger patterns is able to provide new insights.

Sector composition. Analyzing the composition of motifs in terms of which economic sectors are involved, provides interesting insights. If corporate structures were to be organized according to particular motif structures without sectoral preferences, then the division of firms over economic sectors as in Table 2 should be the same for all motif sizes. However, as Figure 4b shows, when the size of motifs

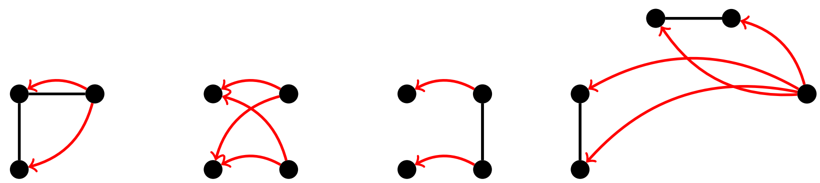


(a) Ratio (horizontal axis) vs concentration (vertical axis) for all patterns. Top right box indicates cut-off values. Patterns of size 3 in blue, size 4 in green and size 5 in red.



(b) Division of firms over economic sectors for the full network and motifs of different sizes.

Fig. 4: Overall (left) and aggregated per motif size (right) results of experiments.



(a) Motif of size 3. (b) Motif of size 4. (c) Motif of size 4. (d) Motif of size 5.

Fig. 5: Four highlighted motifs from the in total 135 discovered motifs, extracted from the $23,048$ enumerated subgraph patterns.

increases, the involvement of the financial sector and banks increases at the cost of a decreasing involvement of the industry sector. This suggests that the financial sector is in general more involved in larger and more complex corporate structures.

7 Conclusion

We have proposed an approach for detecting motifs in multiplex networks by means of adding a layer encoding scheme to a subgraph enumeration algorithm. Through a translation of link types into colored nodes, we integrated the multiplex aspect into the subsequent subgraph labeling step, enabling counting of the frequency of multiplex subgraphs. The resulting approach was tested on a corporate network dataset, and the obtained subgraph frequencies were compared with random networks generated with a multiplex stub-matching model. Experiments showed how multiplex networks provide interesting insight in common business structures. It furthermore highlighted how the financial sector is over-represented in larger constructions.

Whereas this study focused on detecting frequent unweighted multiplex subgraphs, in future work it could be interesting to see the effect of edge weight on the discovered motifs. In corporate networks, this could be used to better distinguish between majority and minority ownership, and its role in different motifs. Finally, it could be interesting to test the algorithm on other multiplex network datasets. In particular, it could be interesting to perform a cross-country comparison, investigating if the prominent presence of particular sectors is prevalent in other countries.

Acknowledgements The first author is supported by funding from the European Research Council (ERC) under the EU Horizon 2020 research and innovation programme (grant agreement 638946). Thanks to the CORPNET group (`http://corpnet.uva.nl`) for useful suggestions.

References

1. Adams, M.: Cross holdings in Germany. Journal of Institutional and Theoretical Economics **155**(1), 80–109 (1999)
2. Alon, U.: Network motifs: Theory and experimental approaches. Nature Reviews Genetics **8**(6), 450–461 (2007)
3. Barabási, A.L.: Network science. Cambridge University Press (2016)
4. Battiston, F., Nicosia, V., Chavez, M., Latora, V.: Multilayer motif analysis of brain networks. Chaos: An Interdisciplinary Journal of Nonlinear Science **27**(4), article 047,404 (2017)
5. Bender, E.A., Canfield, E.R.: The asymptotic number of labeled graphs with given degree sequences. Journal of Combinatorial Theory, Series A **24**(3), 296–307 (1978)
6. Benson, A.R., Gleich, D.F., Leskovec, J.: Higher-order organization of complex networks. Science **353**(6295), 163–166 (2016)
7. Boccaletti, S., Bianconi, G., Criado, R., Del Genio, M., Sendiña-Nadal, I., Wang, Z., Zanin, M.: The structure and dynamics of multilayer networks. Physics Reports **544**(1), 1–122 (2014)
8. Dickison, M.E., Magnani, M., Rossi, L.: Multilayer social networks. Cambridge University Press (2016)
9. Fohlin, C.: The rise of interlocking directorates in imperial Germany. The Economic History Review **52**(2), 307–333 (1999)
10. Garcia-Bernardo, J., Fichtner, J., Takes, F.W., Heemskerk, E.M.: Uncovering offshore financial centers: Conduits and sinks in the global corporate ownership network. Scientific Reports **7**, 6246 (2017)
11. Garcia-Bernardo, J., Takes, F.W.: The effects of data quality on the analysis of corporate board interlock networks. arXiv: 1612.01510 (2017)
12. Ghazizadeh, S., Chawathe, S.S.: SEuS: Structure Extraction Using Summaries. In: Proceedings of the International Conference on Discovery Science, pp. 71–85 (2002)
13. Haiyan, H., Xifeng, Y., Jiawei, H., Jasmine, Z.X.: Mining coherent dense subgraphs across massive biological networks for functional discovery. Bioinformatics **21**(1), 213–221 (2005)
14. Heemskerk, E.M., Takes, F.W.: The corporate elite community structure of global capitalism. New Political Economy **21**(1), 90–118 (2016)
15. Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J.P., Moreno, Y., Porter, M.A.: Multilayer networks. Journal of Complex Networks **2**(3), 203–271 (2014)
16. Märtens, M., Meier, J., Hillebrand, A., Tewarie, P., Van Mieghem, P.: Brain network clustering with information flow motifs. Applied Network Science **2**(1), 25 (2017)
17. McKay, B.D., Piperno, A.: Practical graph isomorphism, II. Journal of Symbolic Computation **60**, 94–112 (2014)
18. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U.: Network motifs: Simple building blocks of complex networks. Science **298**(5594), 824–827 (2002)
19. Mizruchi, M.S.: What do interlocks do? An analysis, critique, and assessment of research on interlocking directorates. Annual review of sociology **22**(1), 271–298 (1996)
20. Ohnishi, T., Takayasu, H., Takayasu, M.: Network motifs in an inter-firm network. Journal of Economic Interaction and Coordination **5**(2), 171–180 (2010)
21. Ribeiro, P., Silva, F.: G-tries: An efficient data structure for discovering network motifs. In: Proceedings of the ACM Symposium on Applied Computing, pp. 1559–1566 (2010)
22. Saeed, S., Saeed, J.: Fast parallel all-subgraph enumeration using multicore machines. Scientific Programming **2015**, 901,321 (2015)
23. Solé-Ribalta, A., De Domenico, M., Arenas, A.: Centrality rankings in multiplex networks. In: Proceedings of the International Conference on Web Science, pp. 149–155 (2014)
24. Takes, F.W., Heemskerk, E.M.: Centrality in the global network of corporate control. Social Network Analysis and Mining **6**(1), 97 (2016)
25. Vitali, S., Glattfelder, J.B., Battiston, S.: The network of global corporate control. PloS one **6**(10), e25,995 (2011)
26. Wernicke, S.: A faster algorithm for detecting network motifs. In: Proceedings of the Workshop on Algorithms in Bioinformatics, pp. 165–177 (2005)