International Workshop on Data Mining on IoT Systems (DaMIS16)

Triadic Motifs in the Partitioned World Trade Web

Antonio Maratea[a,*], Alfredo Petrosino[a], Mario Manzo[b]

[a]*Department of Science and Technologies, University of Naples "Parthenope", Isola C4, Centro Direzionale, 80143 Napoli, Italy*
[b]*Information Technology Services, University of Naples "Parthenope", via F. Acton, 38, 80133 Napoli, Italy*

Abstract

One of the crucial aspects of the Internet of Things that influences the effectiveness of communication among devices is the communication model, for which no universal solution exists. The actual interaction pattern can in general be represented as a directed graph, whose nodes represent the "Things" and whose directed edges represent the sent messages. Frequent patterns can identify channels or infrastructures to be strengthened and can help in choosing the most suitable message routing schema or network protocol. In general, frequent patterns have been called *motifs* and overrepresented motifs have been recognized to be the low-level building blocks of networks and to be useful to explain many of their properties, playing a relevant role in determining their dynamic and evolution. In this paper triadic motifs are found first partitioning a network by strength of connections and then analyzing the partitions separately. The case study is the World Trade Web (WTW), that is the directed graph connecting world Countries with trade relationships, with the aim of finding its topological characterization in terms of motifs and isolating the key factors underlying its evolution. The WTW has been split based on the weights of the graph to highlight structural differences between the big players in terms of volumes of trade and the rest of the world. As test case, the period 2003-2010 has been analyzed, to show the structural effect of the economical crisis in the year 2007.
Keywords: Motif analysis, World Trade Web, subgraph mining, communication models in IoT.

1. Introduction

While any definition is necessarily imprecise, the "Internet of Things" (IoT from now on) commonly refers to objects (Things) of everyday use that have the ability to communicate among themselves and to acquire, share and process data from the surrounding environment. Considering that the IoT enabled devices are rapidly reducing their cost and size, at the same time increasing their computing power, it is easy to forecast an exponential growth in the number of embedded systems able to communicate and in the software applications able to exploit them (see [1] for a survey). Application areas go from the wearable device for online personalized health care, to smart cities or augmented reality and cultural heritage [2]. Necessary conditions for each Thing in the IoT are unique identifiers, high level communication protocols, high level abstractions of the automation possibilities of each device and widespread standards for representing, storing and processing harvested data. One of the crucial aspect of IoT that affects the traffic

* Corresponding author. Tel.: +39-081-5476687 ; fax: +39-081-5476514.
 E-mail address: antonio.maratea@uniparthenope.it

load and the effectiveness of communication among devices is the communication model, for which four different proposal emerged in the literature:

- device to device;
- device to cloud;
- device to gateway;
- back-end data sharing.

Whatever the chosen model is, the actual communication pattern can in general be represented as a directed graph, whose nodes represent the "Things" and whose directed edges represent the sent messages. Interestingly enough, in case of the device to device communication model, things communicate over a network that mimics the topology of social networks or human organizations. In such a scenario, frequent patterns of interaction among nodes can identify channels or infrastructures to be strengthened and can help in choosing the most suitable message routing schema or network protocol. The effectiveness of communication ultimately depends on the network topology and on the chosen protocol.

Given a Network, the frequent patterns of interaction among nodes, that is overrepresented subgraphs of predefined dimensions, are called *motifs*: there are 13 possible motifs on directed graphs with 3 nodes and 199 possible motifs of 4 nodes. Small motifs (size 3 or 4) can be considered as building blocks of complex networks and are useful indicators of the local structure of the network, often helping to explain many of the general properties of the studied system. It has been observed[3] that not all functional motifs playing a role in network dynamics or evolution are statistically overrepresented, but they are nonetheless relevant to highlight local interaction patterns that characterize the network. To evaluate if a motif is statistically significant it is common to compute a *z*-score and the corresponding *p*-value randomizing the network in a way that preserves as much as possible its structure, that is generating many graphs similar to the original and evaluating the empirical distribution of each pattern. Well known problems to be addressed to find motifs are due essentially to the broad search space to be explored, with consequent impractical slowness, especially for higher order subgraphs: subgraph isomorphism is a *NP*-hard problem and the number of subgraphs to be analyzed grows exponentially with graph and motif size. To cope with the broadness of the search space probabilistic algorithms and alternative measures of frequency, allowing or preventing node and edge overlap, can be used. The frequency of a motif is by definition the number of times that the motif is found as subgraph of a given network. While counting matches, nodes and edges overlap can be allowed or prevented, so three different measures of frequency exist: $F1$ that allows overlap of nodes and edges arbitrarily; $F2$ that allows node overlap and prevents edge overlap; $F3$ that prevents any overlap of nodes and edges among subgraphs. To estimate empirically the frequency of each motif it is necessary to generate a number of random networks, so a suitable algorithm for randomization must be provided: the most common is edge switching among nodes, keeping constant the directionality, the input degree and the output degree of each node. A detailed survey on all these facets can be found in[4,5]. Application domains for the motifs search are countless, just to name a few: gene regulatory networks, PPI (protein-protein interactions) graphs and neuronal connectivity graphs in biology and biochemistry; food networks in ecology; electronic circuits and routing algorithms in engineering; the World Trade Web (WTW from now on) in economics. Among others, the WTW is of peculiar interest since it is the graph obtained considering as nodes the Countries of the World Trade Organization and as weighted edges the import/export relationships among Countries, with billions of US dollars as the unity of weight for the total exchange volume. It is a well studied graph, prototype of a complex network, whose topological properties have been explored explicitly since the early 2000 years[6,7,8,9,10,11,12,13,14,15,16,17] and that is known to be scale-free, with a power law distribution of the degrees of the nodes. In the following, the search for overrepresented three nodes patterns in two subgraphs obtained splitting the export WTW based on the strength of the connections will be considered, limited to the period 2003-2010, trying to highlight the structural effect of the economical crisis in the year 2007.

2. Algorithm

In order to count the exact frequency of a network motif of size k it is necessary first to enumerate all subgraphs of size k in the network, then to group them into isomorphic classes. After that, a random graph model is needed

to evaluate if a motif class is overrepresented. The fast RAND-ESU algorithm proposed in[18] is used hereafter: it implements an efficient strategy to enumerate all size-*k* subgraphs building a recursive tree called an ESU-tree for each node in the network. Assuming that all nodes are labeled with the integers $1, \ldots, n$ and that the neighbourhood of a set of nodes $\{C_i\}$ is the set of nodes $\{N_j\}$ adjacent to at least one node in $\{C_i\}$, the algorithm ESU (Enumerate SUbgraps) builds the tree recursively for each node, starting from one node and adding to it all nodes (up to size *k*) that have two properties: the index of the added nodes must be larger than the starting node and the set of new neighbours must exclude the nodes already considered in the neighbours of previous nodes. The obtained tree is proven to have exactly one leaf for each size *k* subgraph. The most important facts about ESU are that the ESU-tree can be efficiently sampled to extract subgraphs without any bias and that its paths can be randomly explored to estimate the total number of size *k* subgraphs. Thanks to the ESU-tree, full enumeration of the subgraphs can be avoided to estimate frequency and an efficient unbiased probabilistic sampling, called RAND-ESU, can be used to randomly skip subgraphs. The idea behind RAND-ESU is to call the recursive function with a given probability. It corresponds to exploring only a section of the ESU-tree, but in a way such that each leaf can be reached with the same probability. RAND-ESU uses the $F1$ frequency.

3. Data

Data are obtained from the UN Comtrade website[1], querying the online system. The metadata used for motif detection are: commodities (all), reporter code (all), partner code (all), year (corresponding to the performed tests), trade flow (import/export). The total trade value in dollars is the weight measuring the strength of connections. Subsequently, the data are filtered by isolating only directional information in the graph (reporter code and partner code) representing the flow of commodities between the different partners. When more than one year is considered, the weights representing flow values in each year have been added to obtain a cumulative value:

$$S_{ij} = \sum_{years} C_{ij}, \tag{1}$$

where S_{ij} is the cumulative flow of commodities in US dollars from Country C_i to Country C_j over the considered years. One graph is obtained representing all the considered years and S_{ij} is the weight of the edges between nodes in this graph. Theoretically if the Country C_i has an export connection with the Country C_j then the latter should have an import connection with the former with the same weight and opposite direction. Due to the differences in reporting procedures and data sources, in practice this equality does not hold: the import graph cannot be obtained simply switching the direction of edges of the export graph and vice versa. While this lack of homogeneity is a generally acknowledged fact, the common practice of averaging the two values does not work for macroscopic differences. In the performed experiments, we have found graphs of import and export of the same year with different number of nodes and edges, together with marked differences in weights very unlikely to be averaged out. Furthermore, it must be observed that in the search for motifs, switching the direction of edges in the network changes also the overrepresented motifs, and there is not a straightforward correspondence between a motif and its reverse. For these reasons, only the directed graph corresponding to export relationships is considered from now on.

4. Motif Analysis of WTW

Topological properties depend by definition upon the structure of the network, and edge weights do not influence them, so that weak connections have the same relevance of strong connections. To highlight how the strength of connections can influence the topology, in terms of correlations or structural differences among strongly and weakly connected nodes, a network split is proposed in this paper. The first step is to split each graph in two subgraphs, called "inliers" and "outliers", based on the edge weights. Table 1 reports threshold values adopted to cut the WTW graphs. Given the highly skewed distribution of weights, there is an evident discontinuity threshold in the histograms, that

---

[1] http://comtrade.un.org

after a number of trials have been found optimal with 10 000 bins. Figure 1 reports the histogram of weights and table 1 reports the chosen thresholds.

Table 1: Thresholds chosen to cut WTW graph into "inliers" and "outliers".

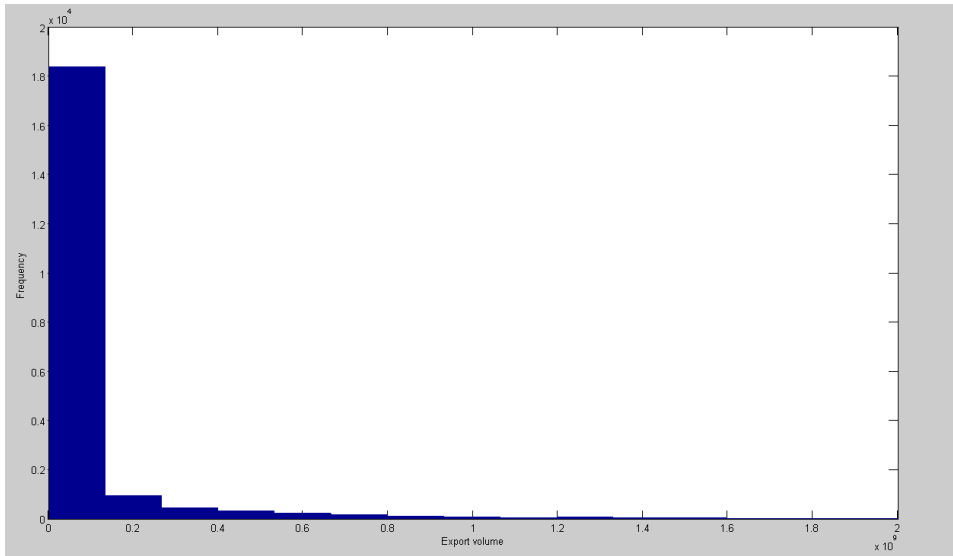| 2004-2006	2007	2008-2010
301 083 696	132 884 136	421 010 358

Fig. 1: Histograms of WTW flow values in US dollars, year 2007.

Analyzing separately the inliers and outliers graphs with RAND-ESU produces Table 2, where the average abundance and the p-value of significant motifs of size 3 on the export WTW from year 2004 to year 2006 are reported.

Table 2: Results on the three-year period 2004-2006. Start year, end year, type, motif ID, percentage of abundance and P-value are reported.

Syear	Eyear	Type	IDM	Percentage	P
2004	2006	Inliers	6	14,15%	0,990
2004	2006	Inliers	12	4,31%	0,000
2004	2006	Inliers	14	21,66%	0,003
2004	2006	Inliers	36	4,40%	0,002
2004	2006	Inliers	46	9,44%	0,020
2004	2006	Inliers	164	2,30%	0,000
2004	2006	Inliers	238	4,64%	0,004
2004	2006	Outliers	6	9,70%	0,000
2004	2006	Outliers	12	4,10%	0,000
2004	2006	Outliers	14	28,62%	0,008
2004	2006	Outliers	36	10,14%	0,999
2004	2006	Outliers	46	4,13%	0,000
2004	2006	Outliers	164	6,02%	0,005
2004	2006	Outliers	238	7,00%	0,000

IDM is the code for the motif (legend in figure 2). The purpose was to highlight the most evident features, so two filters were used: one on the significance (p-value less than 0.05) and another one on frequency (more than 0.05). Results for all motifs that met at least one of these criteria are reported. As test case, data before, after and during the financial crisis of 2007 are shown in tables 2, 3 and 4.

Table 5 gives an overview of the differences before, during and after the 2007 financial crisis. The most evident difference, ca 16%, is between outliers and inliers in motif 164, that after the crisis has a marked increase in abundance. The second major difference, ca 8%, is in motif 14, that is consistently more abundant in outliers than in inliers in all
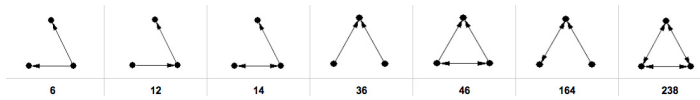
Fig. 2: Significant motifs distinct by id.

Table 3: Results on the year 2007.

Syear	Eyear	Type	IDM	Percentage	P
2007	2007	Inliers	6	15,91%	0,995
2007	2007	Inliers	12	5,60%	0,000
2007	2007	Inliers	14	22,37%	0,000
2007	2007	Inliers	36	4,80%	0,000
2007	2007	Inliers	46	8,14%	0,071
2007	2007	Inliers	164	2,50%	0,000
2007	2007	Inliers	238	2,10%	0,003
2007	2007	Outliers	6	10,95%	0,000
2007	2007	Outliers	12	4,37%	0,000
2007	2007	Outliers	14	29,29%	0,006
2007	2007	Outliers	36	9,20%	1,000
2007	2007	Outliers	46	4,50%	0,000
2007	2007	Outliers	164	5,43%	0,001
2007	2007	Outliers	238	6,88%	0,000

Table 4: Results on the three-year period 2008-2010.

Syear	Eyear	Type	IDM	Percentage	P
2008	2010	Inliers	6	13,64%	0,979
2008	2010	Inliers	12	4,39%	0,000
2008	2010	Inliers	14	21,30%	0,000
2008	2010	Inliers	36	4,29%	0,000
2008	2010	Inliers	46	9,94%	0,036
2008	2010	Inliers	164	21,14%	0,000
2008	2010	Inliers	238	5,51%	0,003
2008	2010	Outliers	6	10,61%	0,000
2008	2010	Outliers	12	4,30%	0,000
2008	2010	Outliers	14	29,69%	0,007
2008	2010	Outliers	36	9,15%	1,000
2008	2010	Outliers	46	4,52%	0,000
2008	2010	Outliers	164	5,04%	0,005
2008	2010	Outliers	238	6,94%	0,000

tested cases. Motif 46 before and after the crisis and motif 238 (fully connected triad) during the crisis are the ones that show the third major difference between inliers and outliers, in the order of 5%. During the year 2007 the motif 46 (uplinked mutual dyad) becomes non significant in the inliers and the difference between outliers and inliers in pattern 238 becomes more evident. As the standard deviations for all computed frequencies are very narrow (in the order of 1/10 000), all differences are statistically significant.

5. Conclusions

Communication models should consider overrepresented motifs to fully exploit the communication potential of a network for a given topology, identifying channels or infrastructures to be strengthened and choosing the most suitable message routing schema or network protocol. Based on the test case of the World Trade Web in the period 2003-2010, differences in abundance and in the number of significant motifs have been highlighted that help to characterize the WTW and to understand the evolution patterns of trade relationship among Countries in terms of volume and time. This represents an hemblematic model for a networks of interactions of human organizations. It has been shown that overrepresented motifs change with the strength of the connections and hence splitting a network is a viable way to gain insight into its structure and dynamics. Future work is in studying motifs of size greater than 3 and the induced effect of self-organization.

Table 5: Results on the three-year period 2008-2010.

IDM	Type	Before	During	After
6	Inliers	–	–	–
6	Outliers	9,70%	10,95%	10,61%
12	Inliers	4,31%	5,60%	4,39%
12	Outliers	4,10%	4,37%	4,30%
	Difference:	**0,21%**	**1,23%**	**0,09%**
14	Inliers	21,66%	22,37%	21,30%
14	Outliers	28,62%	29,29%	29,69%
	Difference:	**-6,96%**	**-6,92%**	**-8,39%**
36	Inliers	4,40%	4,80%	4,29%
36	Outliers	–	–	–
46	Inliers	9,44%	–	9,94%
46	Outliers	4,13%	4,50%	4,52%
	Difference:	**5,31%**	**–**	**5,42%**
164	Inliers	2,30%	2,50%	21,14%
164	Outliers	6,02%	5,43%	5,04%
	Difference:	**-3,72%**	**-2,93%**	**16,10%**
238	Inliers	4,64%	2,10%	5,51%
238	Outliers	7,00%	6,88%	6,94%
	Difference:	**-2,36%**	**-4,78%**	**-1,43%**

References

1. Atzori, L., Iera, A., Morabito, G.. The internet of things: A survey. *Computer Networks* 2010; **54**(15):2787 – 2805. URL: `http://www.sciencedirect.com/science/article/pii/S1389128610001568`. doi:http://dx.doi.org/10.1016/j.comnet.2010.05.010.
2. Cuomo, S., De Michele, P., Galletti, A., Piccialli, F.. A cultural heritage case study of visitor experiences shared on a social network. In: *10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing, 3PGCIC 2015, Krakow, Poland, November 4-6, 2015*. ISBN 978-1-4673-8317-2; 2015, p. 539–544. URL: `http://dx.doi.org/10.1109/3PGCIC.2015.55`. doi:10.1109/3PGCIC.2015.55; publisher: IEEE.
3. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U.. Network motifs: simple building blocks of complex networks. *Science* 2002;**298**(5594):824–827.
4. Kim, W., Diko, M., Rawson, K.. Network motif detection: Algorithms, parallel and cloud computing, and related tools. *Tsinghua Science and Technology* 2013;**18**(5):469–489.
5. Wong, E.A., Baur, B.. On network tools for network motif finding: a survey study. 2010.
6. Chiarucci, R., Ruzzenenti, F., Loffredo, M.I.. Detecting spatial homogeneity in the world trade web with detrended fluctuation analysis. *Physica A: Statistical Mechanics and its Applications* 2014;**401**:1–7.
7. Squartini, T., Garlaschelli, D.. Triadic motifs and dyadic self-organization in the world trade network. In: *Self-Organizing Systems*. Springer; 2012, p. 24–35.
8. Saracco, F., Di Clemente, R., Gabrielli, A., Squartini, T.. Detecting the bipartite world trade web evolution across 2007: a motifs-based analysis. *arXiv preprint arXiv:150803533* 2015;.
9. Fagiolo, G., Reyes, J., Schiavo, S.. The evolution of the world trade web: a weighted-network analysis. *Journal of Evolutionary Economics* 2010;**20**(4):479–514.
10. Li, X., Jin, Y.Y., Chen, G.. Complexity and synchronization of the world trade web. *Physica A: Statistical Mechanics and its Applications* 2003;**328**(1):287–296.
11. Barabási, A.L., Albert, R., Jeong, H.. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A: Statistical Mechanics and its Applications* 2000;**281**(1):69–77.
12. Squartini, T., Fagiolo, G., Garlaschelli, D.. Randomizing world trade. i. a binary network analysis. *Physical Review E* 2011;**84**(4):046117.
13. Ruzzenenti, F., Picciolo, F., Basosi, R., Garlaschelli, D.. Spatial effects in real networks: Measures, null models, and applications. *Physical Review E* 2012;**86**(6):066110.
14. Serrano, M.Á., Boguñá, M., Vespignani, A.. Patterns of dominant flows in the world trade web. *Journal of Economic Interaction and Coordination* 2007;**2**(2):111–124.
15. Garlaschelli, D., Di Matteo, T., Aste, T., Caldarelli, G., Loffredo, M.I.. Interplay between topology and dynamics in the world trade web. *The European Physical Journal B* 2007;**57**(2):159–164.
16. Serrano, M.Á., Boguñá, M.. Topology of the world trade web. *Physical Review E* 2003;**68**(1):015101.
17. Skowron, P., Karpiarz, M., Fronczak, A., Fronczak, P.. Spanning trees of the world trade web: real-world data and the gravity model of trade. *arXiv preprint arXiv:14121618* 2014;.
18. Wernicke, S.. A faster algorithm for detecting network motifs. In: *Algorithms in Bioinformatics*. Springer; 2005, p. 165–177.