



HAL
open science

Time considerations for the study of complex maritime networks

Frédéric Guinand, Yoann Pigné

► **To cite this version:**

Frédéric Guinand, Yoann Pigné. Time considerations for the study of complex maritime networks. *Maritime Networks: Spatial structures and time dynamics*, 1, Routledge, pp.163-189, 2015. <hal-02122859>

HAL Id: hal-02122859

<https://hal.science/hal-02122859v1>

Submitted on 7 May 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Time Considerations for the Study of Complex Maritime Networks

Frédéric Guinand and Yoann Pigné
LITIS - Université du Havre
25 rue Philippe Lebon - 76600 Le Havre
email:firstname.lastname@univ-lehavre.fr

April 8, 2015

1 Introduction

The globalization of exchanges coupled with containerization have considerably increased maritime traffic these last decades. Thanks to the standardization of containers, goods can be transported on long distances by container-ships, trains and trucks, without any change in the conditioning. Containers represent universal transportation boxes that ease the mechanization of the handling system, thus, today, the main part of world trade is carried by sea and most of the goods are carried by container ships.

This globalization has been accompanied by rapid changes in port infrastructures, in shipping companies organization and consequently in the global shipping network. The strategy of major companies has led to a multilevel structure of the network. At the higher level, this global network is now composed of multi-port calling networks (kind of mesh between main ports, linked by mega-containerships) that coexist with hub-and-spoke networks (star networks at the regional level).

Motivated by these changes an increasing amount of works have been conducted on various subjects related to this evolution. Network optimization problems have been widely studied [27] in particular problems of network design, route optimization, or hub location combined or not with the problem of allocation of nodes to hubs [23, 1, 15, 11]. The analysis and the study of the evolution of the shipping network, at worldwide or regional scales, have also received much attention [9, 4, 30, 20, 24, 28].

For a majority of these researches, in complement to statistical analyses, tools and metrics from graph theory have been used. Such an approach is not a novelty [18, 10, 16], however, with the popularization of complex networks theory [31, 22], studies resorting to these methods have multiplied.

In many of these works some properties come back in a recurrent way. Thus, according to the considered periods and data, it has been proved that the distribution of port degrees follows a power law or is an exponential-like distribution. Small world property and the scale free nature of the shipping network is also often mentioned. In addition to these properties, many works also propose some interpretations from the values of centralities, clustering coefficients, average nearest neighbours degrees, and many other graph metrics [17]. From these measures, they draw some conclusions about, the performances of ports [32], changes in their hierarchy [3], evolution of the topological structure of the network [28], that may have several complementary causes: the hub-and-spoke strategies of ports and carriers that leads to a rich-club phenomenon [14, 5, 4], or the world trade situation [20, 24].

In all these works, the studied graphs are obtained from all ship's movements. Indeed, unlike continental transport networks, where the traffic paths rely on existing and real infrastructures (roads, channels, railways, waterways), the maritime network is not characterized by physical links [18], but by links corresponding to "the reality of regular lines defined by container ships' movements" [16]

and "information about the itineraries of [...] cargo ships [are used] to construct a network of links between ports" [17]. These movements can be obtained from various sources like databases [14], web sites of major shipping companies [32], annual reports (like the Containerisation International Yearbooks) [28], Automatic Identification System (AIS) [17] (AIS is a radio-based system located on ships, that provides information about the position, the course and the speed of the ship, however its usage is limited to the transmission range of the radio (less than 30 nautical miles)) or Lloyd's reports [4]. The models and results presented in this chapter were obtained using data coming from these reports: the Lloyd's List Periodicals. The Lloyd's company collects regularly (on a daily basis) information about ships and more than 90% of container ships are tracked by the company. These registers, called Lloyd's List Periodicals (LLP), report many information about ports-ships events like the date at which a ship enters a port, the date at which it leaves this port, its next destination and its previous port.

In most papers from literature, graphs representing shipping networks are built from all ships' movements for the whole considered period (usually one year), whatever their number between two given ports and whatever the dates of departures from these ports. This may be explained by the regularity of maritime routes used by the main carriers. Thus we can say that the *temporal granularity* of the studied graphs corresponds to a coarse grain. However, if the quality and the accuracy of the data allows it, it is possible to consider a finer temporal granularity. In the sequel we propose to take into account this granularity in two different ways, leading to two different graph models.

Instead of building the graph from all the events occurring during one year, in our first model, we propose to build graphs from events occurring during a given fixed period of time. This period of time corresponds to a temporal granularity J , where J is the number of consecutive days of a time-window. The graph, called a TG-Graph, results from the events occurring during this time window. We then build the time series of such graphs of temporal granularity J for the whole year ($365 - J$ graphs). From these series we compute and study a set of simple metrics like the number of nodes, the average degree, the size of the giant connected component and some other measures detailed later in the chapter. We have noticed that during the year some metrics present particular shapes and are sometimes linked, opening the way to some interpretations.

For the second model we have considered, given the data, the smallest temporal granularity. We propose the use of a model able to keep in its own structure all the events changing its topology. Temporal networks [13] and evolving graphs [8] are two examples of families of graphs allowing such a representation. We will see that keeping all these events in the structure allows an analysis of the shipping networks using the container point of view.

The rest of the chapter is made of three main parts. In the first next section, a complete analysis of the commonly used model of graph is performed. In section 3, a new model is introduced for studying the evolution of the shipping networks based on the analysis of time series of TG-Graphs. In the last main section Temporal Networks are described and resulting shipping networks analysed. A conclusion and graphics, gathered in the appendix, end this chapter.

2 Static Graphs

From the data several static graph models can be derived. Nodes are always associated to ports and only the links have a different semantic. Some models are built from maritime routes defined by main carriers companies. A maritime route is an oriented cycle, starting from one port A , to a destination port B with several calls along the way, and coming back to the original port A .

A first graph model can be built in which the arcs correspond to the travel of the ships from one port to the next one on the route. This model corresponds to the space L topology described in [25] or to the GDL (graph of direct links) mentionned in [4]. This graph is generally the model considered in the majority of the study about the evolution of maritime networks. The Global Cargo Ship Network (GCSN) described in [17] is built according to the same principles but without being restricted to container ships.

| Measure | Year 1996 | Year 2006 |
|-----------------------------------|-----------|-----------|
| Nodes | 872 | 1137 |
| Arcs | 7778 | 13704 |
| Connected Components | 2 | 2 |
| Connected Components > 1 | 2 | 2 |
| Giant Component Size | 869 | 1134 |
| Giant Component Ratio | 99.66% | 99.74% |
| Strongly Connected Components | 36 | 30 |
| Strongly Connected Components > 1 | 2 | 2 |
| Giant Strongly Component Size | 835 | 1106 |
| Giant Strongly Component Ratio | 95.76% | 97.27% |
| Average Degree | 17.84 | 24.11 |
| Average Clustering Coefficient | 0.53 | 0.53 |
| Diameter | 7 | 7 |
| Radius | 4 | 4 |
| Path Length | 3.22 | 3.06 |
| Average Eccentricity | 5.6018 | 5.3201 |

Table 1: Common Graph Measures on the two static networks (year 1996 and 2006)

From the same set of data and still based on the notion of maritime route, another graph can be built by joining all the ports belonging to the same route. This corresponds to the space P topology [25] or to the GAL (graph of all links) [4]. This second graph model highlights the connectivity between ports relying on ships movements. A path requiring more than one link for reaching port B from port A implies that transshipments are needed.

If there exist some studies that take into consideration both models [25, 32, 14], most of analyses of shipping networks emphasize the space L topology or GDL model. In this section we focus on the analysis of such a graph built from the data extracted from 1996 and 2006 Lloyd’s List Periodicals. In the obtained graphs (one per year), one node represents one port and the set of nodes represents all ports that were active at least once during the year (1996 or 2006). A port is said to be active if a ship left or arrived in that port during the year. One arc is added between two nodes s (a source node) and d (a destination node) if and only if there exists in the data at least one record indicating that one ship made a direct journey from s to d .

Different kinds of analysis can be carried out on this static network, at different scales. The network can be measured as a whole, with global measures that would give a general overview of the dataset or at a lower scale, at the node level for instance. Global measures usually are scalar values like, the average degree of the network, its radius or its diameter (Table 1 gathers some global measures for the two datasets (year 1996 and 2006)), while the degree of all nodes, or their clustering coefficients are local measures providing vectors of values often analysed as distributions.

The focus is first given to global measures with the study of *Connected Components* and of the *Small World Property*. Then node-level measures are investigated with the *Scale-Free Property*.

2.1 Connected Components

A connected component is a subset of the network where each node of the set is somehow connected (directly or indirectly) to all the other nodes of the subset. The direction of arcs is not considered in connected components. In other words, there exists an undirected path between any pair of nodes in that subset. The more restrictive type of subset that are *strongly connected components* impose that any two nodes in the subset are somehow connected through a directed path (following actual arcs directions). Table 1 advertises two connected components of respective sizes 869 and 3 in 1996 (respectively 1134 and 3 in 2006). This 3-nodes component is identical in both data sets and will be ignored in the remaining. Apart from those 3 ports, the network is connected.

When it comes to strongly connected components, in 1996, 36 components are identified but only

| Year | L | L_{rand} | C | C_{latt} | ω |
|------|-------|------------|-------|------------|----------|
| 1996 | 3.226 | 2.662 | 0.527 | 0.700 | 0.0716 |
| 2006 | 3.062 | 2.558 | 0.527 | 0.717 | 0.1017 |

Table 2: Comparison with random and regular networks in order to point out small-world properties.

two are larger than 1 (components of size equal to 1 are isolated nodes). One of the 2 components is again the 3-nodes component. This leaves 34 nodes out of the giant strongly connected component. These nodes are *leaves* in the network: whether no path can reach them (only path initiated from them is possible) or the opposite (they can be reached but have no departure). The same observation can be done on the year 2006 dataset with 1106 strongly connected nodes and 28 leftovers.

2.2 Small World Property

When considering network path lengths (number of hops between nodes), in both networks, each node's most distant destination ranges between the network radius and its diameter (4 and 7) and the average path length for any pair of nodes is 3.22 in year 1996 (resp. 3.06 in 2006). Moreover the average clustering coefficient for both networks is relatively high. These measures suggest that a *small-world* property [31] should be investigated. Following the ω test [26] we compare the clustering coefficient of our datasets to the one of an *equivalent lattice network* (a regular network with same number of nodes and same average degree). We also compare the path length to an *equivalent random network* (a network with same number of nodes and same average degree as our network [7]). Let L be the path length of our dataset and C its clustering coefficient, L_{rand} is the path length of the random network and C_{latt} is the clustering coefficient of the lattice. Then the *small-world* measurement ω proposed by Telesford *et al.* is as such:

$$\omega = \frac{L_{rand}}{L} - \frac{C}{C_{latt}}$$

Measures indicate an ω value of 0.0716 for year 1996 (resp. 0.1017 for year 2006). See Table 2 for details. Values close to zero are the witnesses of the small-world property. Positive values tend to indicate that the networks have more random characteristics than regular ones. From these results we conclude with $\omega \approx 0.1$, $L > L_{rand}$, and $C < C_{latt}$, that the 1996 and 2006 maritime networks present a small-world property but with slightly more random characteristics due to a lower clustering coefficient than the lattice equivalent graph.

2.3 Scale-Free Property

Additional information can be obtained from local measures when analysing the distributions of their values. Networks with degree distributions following a power law are classically said to be Scale-Free networks. Verifying such a property for maritime networks would help classify them and understand their underlying mechanisms. However, deciding that a network is Scale-Free requires careful analysis and comparisons. Basically Scale-Free networks have a degree distribution that follows a power law.

Plotting the degree distribution of a supposed Scale-Free network is however not very informative. Indeed, since very few nodes in the network gather the highest degrees, the visual representation at the right of the distribution gets statistically noisy as fewer samples express more values. This phenomenon, also called *fat tail* prevents precise observation. It is thus desirable not to draw the distribution (its *probability density function*) itself but its *cumulative distribution function* (CDF) which also follows a power law if the original distribution follows one [21].

It is also rare that the degree distribution of the network can be identified as a power law on the full range of the distribution. It is admitted [2] that there exists a threshold value x_{min} from which a distribution obeys a power law.

If a power law can be matched to a real dataset, it is also desirable to try to match it against other random distributions that might fit the data better.

In this chapter guidelines from Clauset *et al.* [2] have been followed in order to analyse and fit our two datasets. This method follows three main steps.

1. Estimate the x_{min} threshold and the α parameter of the power law that would fit the real data. The power law parameter, α , is estimated with the method of *maximum likelihood* (MLE). The lower bound x_{min} is estimated by minimizing the distance between the CDF and the data using the Kolmogorov–Smirnov (KS) distance.
2. Quantitatively estimate how good the fit is between the data and the power law. This method generates synthetic data based on α and tries to recover the x_{min} lower bound with the MLE. The result is mainly a standard deviation for the given value of x_{min} . The lower the deviation, the better the fit. Also, the KS distance is used to estimate if those synthetic generated data are closer or not to the theoretical power law than the data itself is. If the data outperforms at least 10% of the synthetic data (its p-value) then it is considered a plausible match for that power law.
3. Compare the power law with other distributions. Indeed other random distributions might fit the data. The two first steps can be applied to evaluate them. Finally a likelihood test ratio can give a pairwise comparison between distributions.

Figure 1 displays the CDF of the degree distribution of the 1996 dataset along with estimates for a power-law, a Log-normal, and an exponential distribution. Each distribution is drawn with its own estimated value for the lower bound x_{min} . These results and the subsequent tables and figures were all generated using the R package *powerLaw* [12]. Distribution estimates for the year 2006 dataset, not shown here, advertise comparable behaviour.

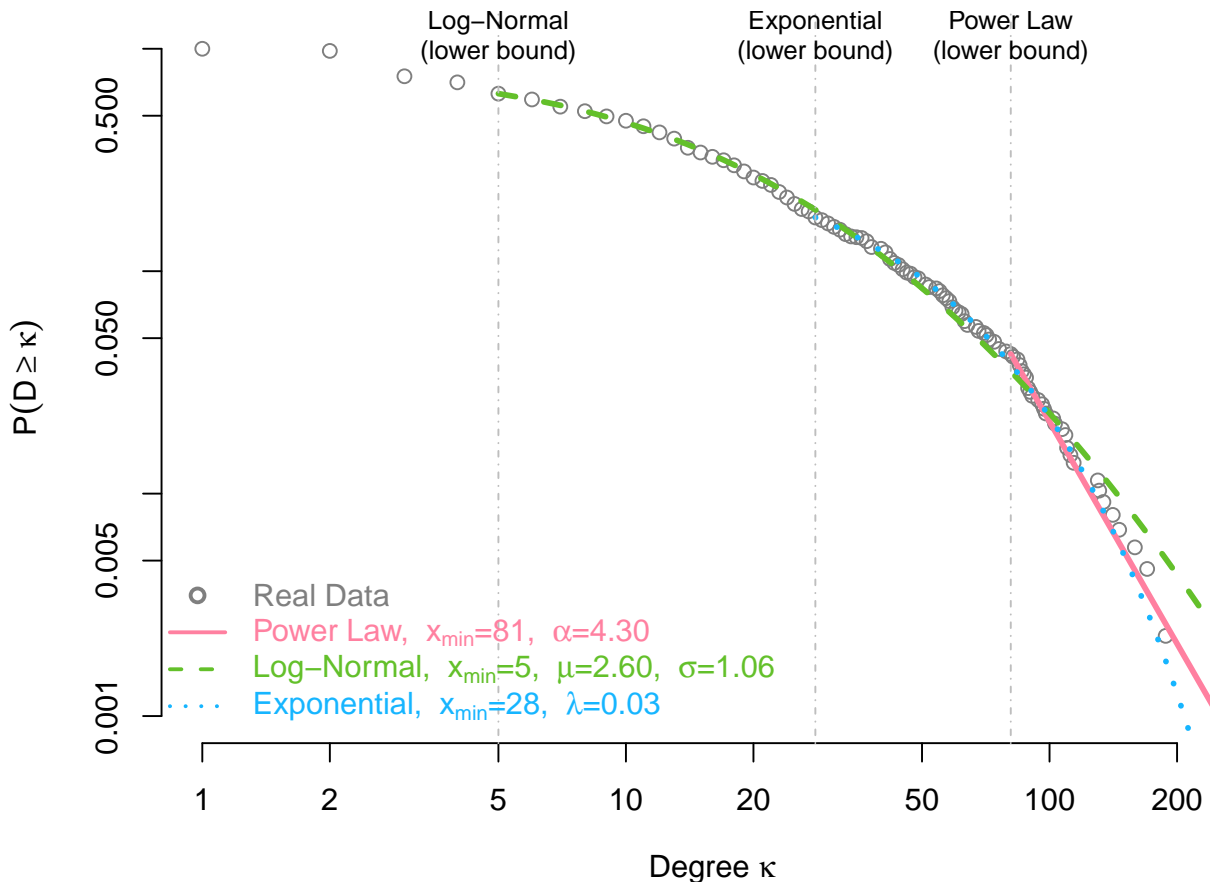


Figure 1: Degree Distribution's cumulative density function estimated with various discrete probability distributions (year 1996). Each distribution has fitted parameters estimated with the method of *maximum likelihood* and lower bounds x_{min} optimised with the KS distance method.

| Year | Law | x_{min} | n_{tail} | p-value |
|------|-------------|-------------|---------------|---------|
| 1996 | Power Law | 81 ± 20 | 37 ± 37 | 0.89 |
| | Log-normal | 5 ± 5 | 548 ± 458 | 0.53 |
| | Exponential | 28 ± 10 | 152 ± 133 | 0.94 |
| 2006 | Power Law | 71 ± 22 | 102 ± 108 | 0.32 |
| | Log-normal | 27 ± 11 | 295 ± 196 | 0.67 |
| | Exponential | 53 ± 17 | 145 ± 127 | 0.52 |

Table 3: Various discrete probability distributions fits with p-values for the two datasets. x_{min} is the estimated lower bound on the considered law (with standard deviation). n_{tail} is the quantity of observations involved (with standard deviation). The p-value is the ratio of comparisons between the real data and synthetic datasets that gives the real data a better fit for the considered distribution. P-values above 0.1 are considered a plausible match for the considered distribution.

The estimation of parameters visually shows that the x_{min} estimated for the power law is very high, making the distribution fit only a fraction of the dataset, while the log-normal distribution has a lower x_{min} estimated which fits more data. The quantitative estimate of those parameters with a bootstrapping methods (Table 3) shows that not only the power law has a very high x_{min} estimate, making it fit less real observations (n_{tail}), but the estimate is statistical unstable with high standard deviation, meaning that the x_{min} value is barely estimated on synthetic datasets with same characteristics. The log-normal distribution, on the other hand behaves well on both datasets fitting the largest quantity of observation.

P-values, that indicate how close the data is to the distributions compared to generated synthetic datasets, are all above the 0.1 threshold which indicates that the three distribution, independently, are a *plausible* fit to the dataset.

Distributions are then compared one against the other. The sign of the *likelihood ratio test* [2] indicates which distribution is the best fit for the dataset and Vuong’s test indicates the reliability of the *likelihood ratio test* based on its standard deviation. In order to test distribution together, the lower bound x_{min} must be equal for both. Figure 2 shows pairwise comparison between the three considered laws depending on the value of x_{min} for the 1996 dataset. The power law fit is always ruled out by the two other distributions, especially for small values of the x_{min} lower bound. The log-normal distribution overpasses the exponential however the Vuong’s test gives no credit to that last assumption. Comparable comments can be made on the year 2006 dataset (not shown here).

A first conclusion could be that the MLE estimation for a power law to fit the dataset is not ruled out. There exists a power law for the data and we can qualify the networks as Scale-Free. However, further analysis show that other distributions, especially log-normal are a better candidate to fit these data. For more comparison with sub-exponential distributions, and comparable datasets, the reader is invited to refer to Gastner and Ducruet in this same book.

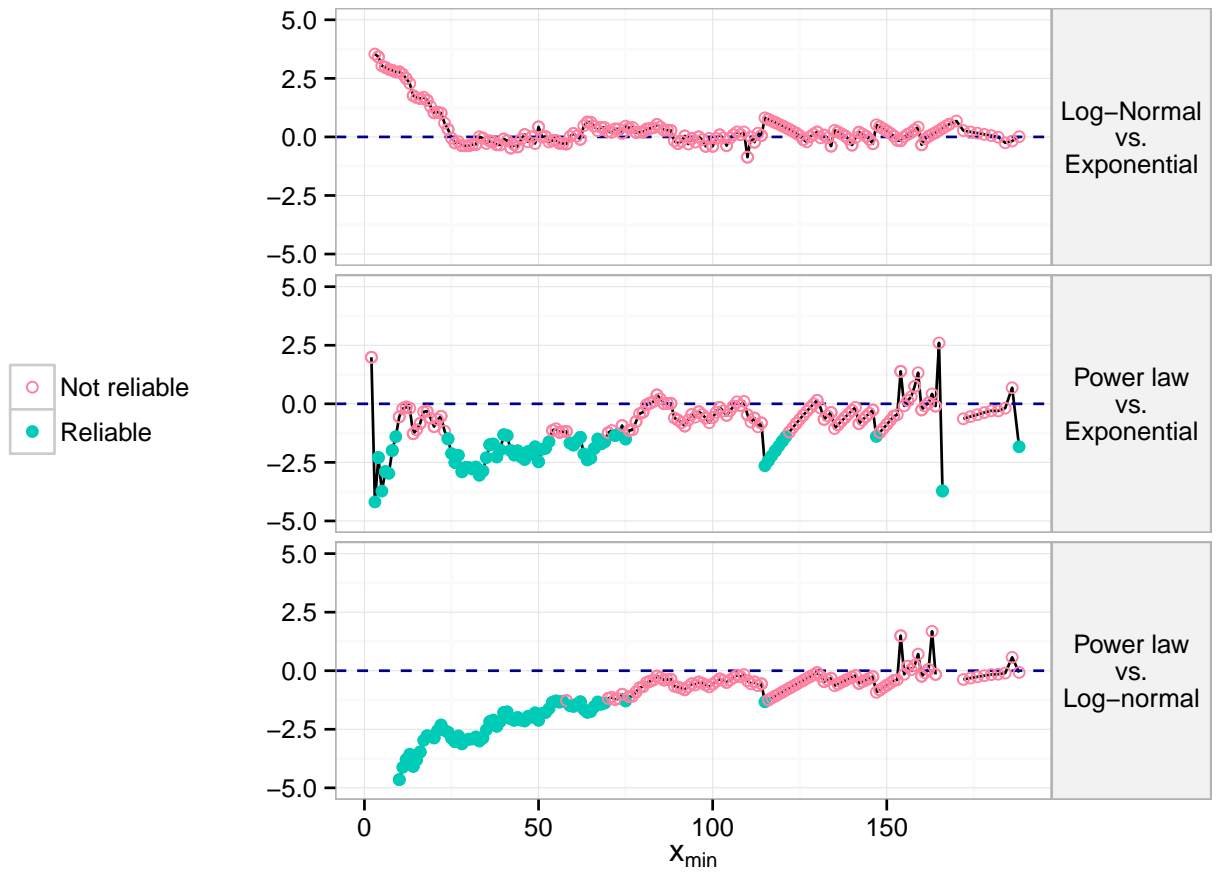


Figure 2: Distributions comparison with *likelihood ratio test* and estimation of reliability with Vuong's test. The power law fit is always ruled out by the two other distributions, especially for small values of the x_{min} lower bound. The Vuong's tests does not allow any conclusion about the log-normal vs. exponential comparison.

3 Graphs based on Time Granularity: TG-Graphs

Graph models described in the previous section are built from the whole set of ships' movements recorded on a long time period (usually one full year). However, if the quality and the accuracy of the data allows it, it is possible to consider a smaller time granularity and to build graphs from a restricted set of ships' movements. This restriction can be spatial (restriction to a given region) or temporal (restriction to a given time period). The latter interests us. The time window considered can be of one day, up to several months. A TG-Graph (Time Granularity-based Graph) is defined by a starting date t and a time granularity J . t and J define a time window $W_{t,J}$ of length J (days) and starting at t . The TG-Graph $TGG(t, J)$ is built from all ships' departures (and not arrivals) occurring during $W_{t,J}$. An arc (s, d) is added to the graph if and only if one ship, at least, left the port s to port d at a date belonging to the time interval $[t, t + J - 1]$ (or equivalently during $W_{t,J}$). Nodes are added to the graph accordingly: a node (port) p belongs to $TGG(t, J)$ if and only if there exists a ship departure in $W_{t,J}$ for which p is either the source port or the destination port.

Let us consider the set of departures given in Table below. $TGG(1, 3)$ and $TGG(3, 3)$ are represented in Figure 3.

| | | | | | | | | | |
|------------------|---|---|---|---|---|---|---|---|---|
| Source port | A | B | B | B | C | C | C | D | E |
| Destination port | B | D | E | E | B | D | D | C | D |
| Departure date | 1 | 5 | 2 | 4 | 1 | 2 | 4 | 4 | 3 |

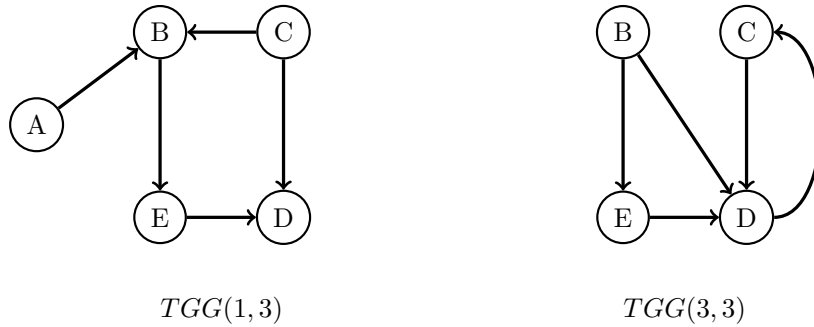


Figure 3: $TGG(t, J)$ are graphs built from ships' departures occurring during the time window $W(t, J)$, or equivalently during the time interval $[t, t+J-1]$. On the left side of the figure, $TGG(1,3)$ is the TG-Graph built from ships' departures occurring at dates 1, 2 and 3. On the right side, $TGG(3,3)$ is the graph built from ships' departures occurring at dates 3, 4 and 5.

3.1 Time Granularity and measures

Given a time granularity J we can build $365 - J$ TG-graphs for 2006 and $366 - J$ for 1996, with starting dates from $t = 1$ to $t = 365 - J$ ($t = 366 - J$ for 1996), leading to time series of TG-graphs. For each TG-graph we can determine classical graph measures: graph order, diameter, radius, eccentricity, average degree, giant component size, and many other classical metrics and, from the time series, we can study their evolution along the year, looking for tendencies or particular shapes of the graphics. We denote by $TGG(J, 1996)$ the series of TG-Graphs with granularity J for year 1996 for starting dates from $t = 1$ to $t = 366 - J$.

In this work we consider non-directed TG-Graphs and we focus mainly on the graph order (number of nodes), the average node degree, the size of the giant connected component, the average eccentricity and the path length. But, for studying time series of TG-Graphs we first have to choose time granularities. How to choose a relevant set of time granularities is an open question, thus, we have chosen, empirically, the following set of values: 1, 7, 14, 21 and 28 days. All graphics obtained for the different measures are reported in appendix (the giant strongly connected component ratio was the only measure obtained from directed TG-Graphs).

3.2 Observations and Analyses

One of the most interesting measures of the obtained results is the evolution of the number of nodes for different values of time granularity. For $TGG(1, 1996)$ we can see on Figure 4 a *Sunday effect*, that plays an important role in the variability of the number of nodes in the network on a daily basis.

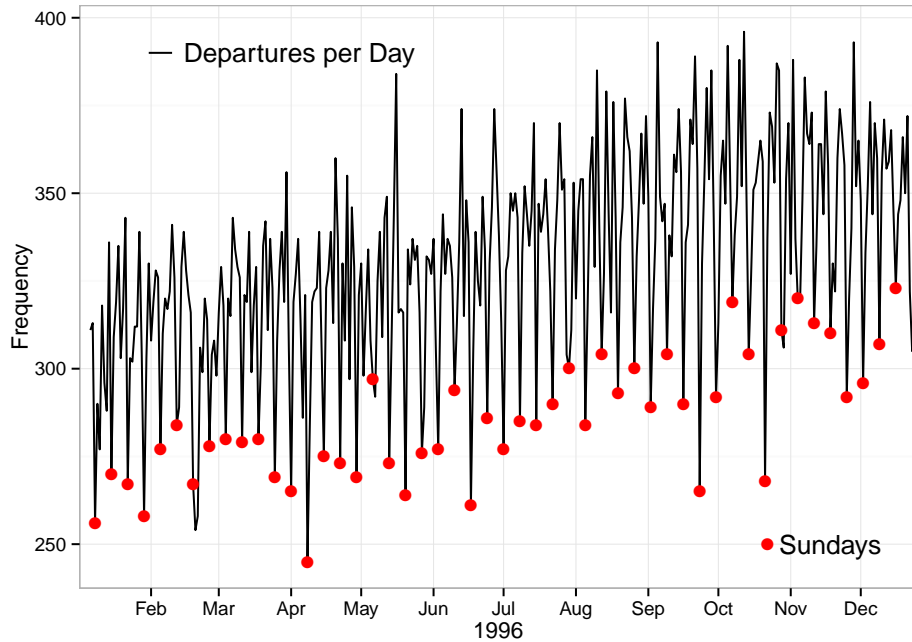


Figure 4: Day-per-day departures frequencies for all ports during year 1996. Dots indicate Sundays.

We can see that between 1996 and 2006 this effect has been levelled as illustrated by Figure 5, but we leave the interpretation of this effect to geographers or sociologists.

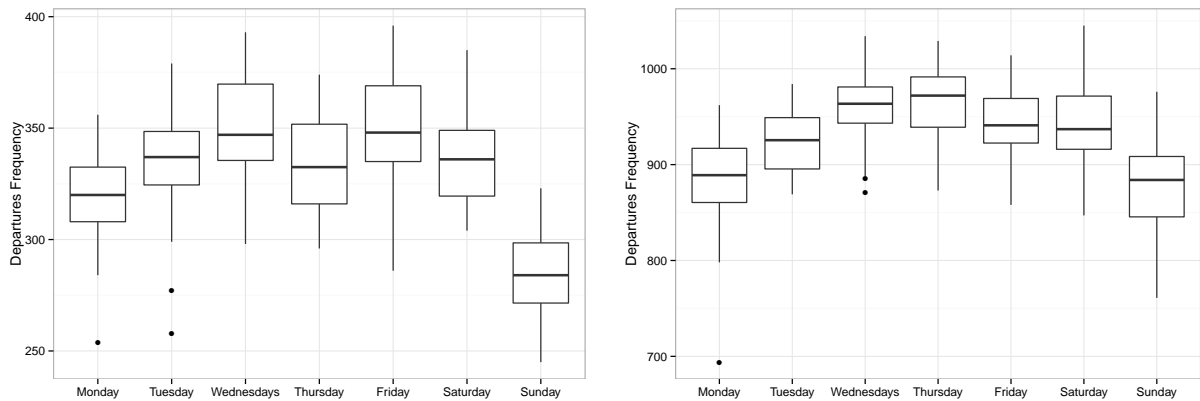


Figure 5: Number of departures per day in 1996 (left) and in 2006 (right)

At this level of granularity ($J = 1$) the signal is noisy, however it is already possible to notice a tendency for the whole year, an increase in the number of nodes involved in maritime traffic. This tendency is verified for both 1996 and 2006. If we consider the growth rate in 1996, obtained by linear regression, and if we assume that it remained unchanged between 1996 and 2006, the number of nodes at the end of 2006 would have been about 480. The actual number is 350. This suggests a slowdown in the number of ports involved in the shipping network (on a daily basis) between 1996 and 2006. This is verified, at least for 2006, since the value of the growth rate in 2006 is more than three times smaller than the one of 1996, as illustrated by Figure 6. If the regularity of the slowdown between 1996 and 2006 could be proved, which might be confirmed by additional data for

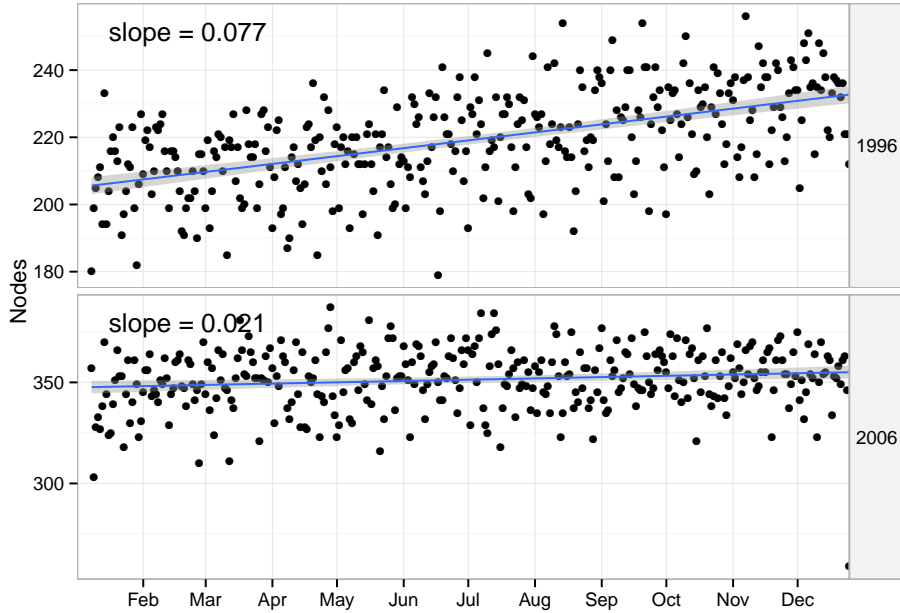


Figure 6: Rate of growth of the number of nodes for time granularity $J = 1$. In 1996 (up) and 2006 (down).

years between 1996 and 2006 and after 2006, this would suggest that the shipping network would reach (quite quickly) a threshold on the number of ports involved in container shipping. Then, the main evolution process of the network would be an enhancement of the connectivity of ports.

Most often if we increase the value of time granularity the impact on the graphics is evident. However, the *sensitivity* of the metric to time granularity is not the same for all. Often, increasing time granularity reduces the variability of the measures' values and also changes those values. For instance the number of nodes in $TGG(1, 1996)$ belongs to the interval $[200, 240]$ while this interval is equal to $[490, 540]$ for $TGG(28, 1996)$. Moreover, if we look at the graphics presented in Figure 7, some structures emerge in the graphics when time granularity increases. Patterns appear and become more visible when sequentially observing time series of nodes' evolution for graphs $TGG(1, 1996)$, $TGG(14, 1996)$, $TGG(21, 1996)$ and $TGG(28, 1996)$ (the interpretation of such patterns is still under investigation).

However, such a sensitivity is not a general rule. Indeed, the average eccentricity value belongs to the interval $[6, 8]$ for the vast majority of the measures for $J \geq 7$ in 2006 or $J \geq 14$ in 1996. We can also notice that for $J = 1$ this value is much larger in 2006, as it is the case for $J = 1$ and $J = 7$ in 1996. While a deeper analysis is still to be done, this could indicate that, for some metrics, there exists a threshold value, for time granularity, from which the graph becomes stable for this measure. For the average eccentricity this stability is reached in 2006 for lower values of J than in 1996, what could be interpreted as an increase in the frequency of the maritime traffic.

Another way of analyzing time series of TG-Graphs is to compare the evolution of couples of measures. Such comparisons can instigate some hypotheses on the corresponding graphs that could help understanding shipping networks evolution. The most illustrative example for our data is given by the compared evolution of the number of nodes and the average node degrees, for both years and especially for $J = 28$. Figures 8 and 9 highlight this link.

Both graphics suggest that when the number of nodes increases the average degree decreases. From a graph perspective this could indicate that all TG-Graphs have in common a set of always present nodes characterized by a high degree. During the year, some low degree nodes are linked to this common structure and leave it later. From the maritime network perspective, this could be interpreted as a manifestation of the hub-and-spoke structuring of the shipping network.

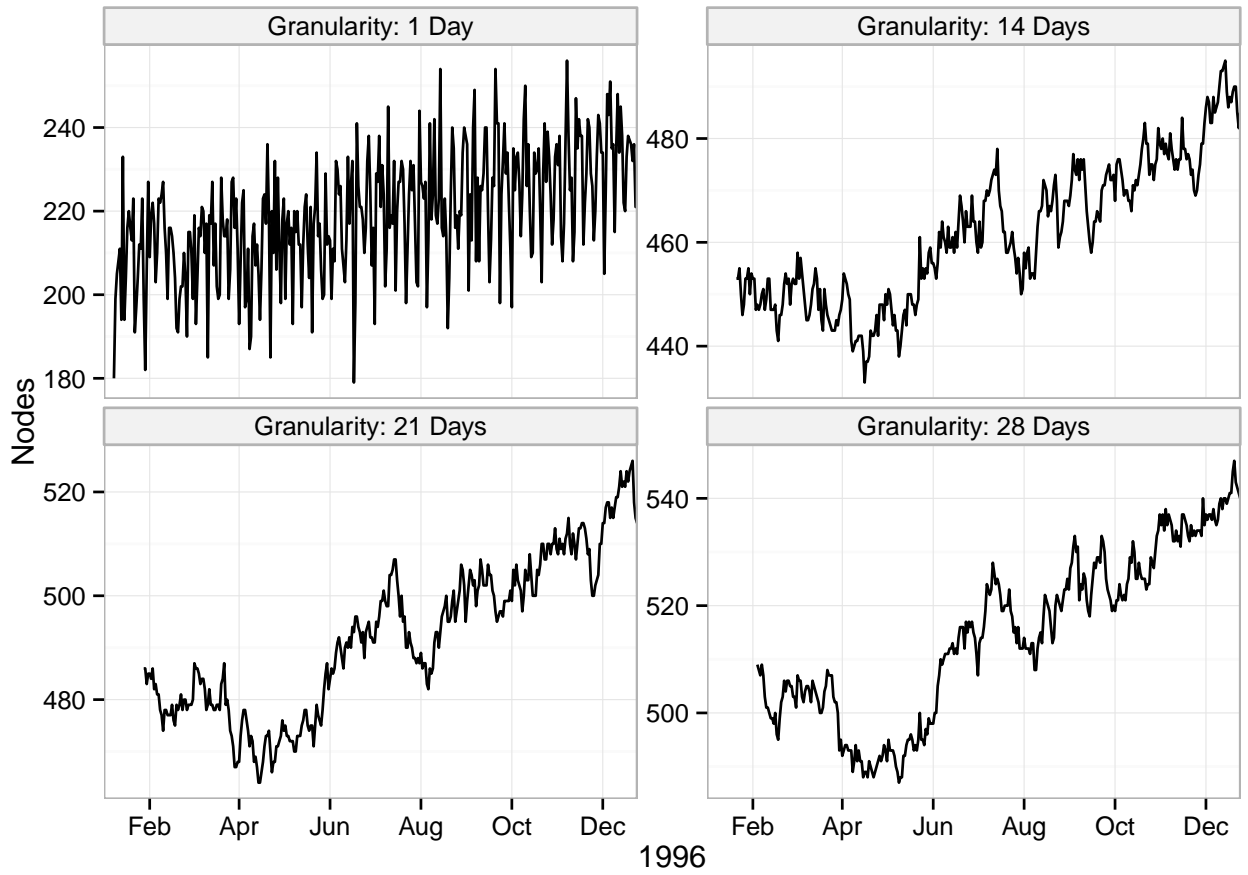


Figure 7: Evolution of the number of nodes in TG-Graphs according to the value of J (from 1 to 28)

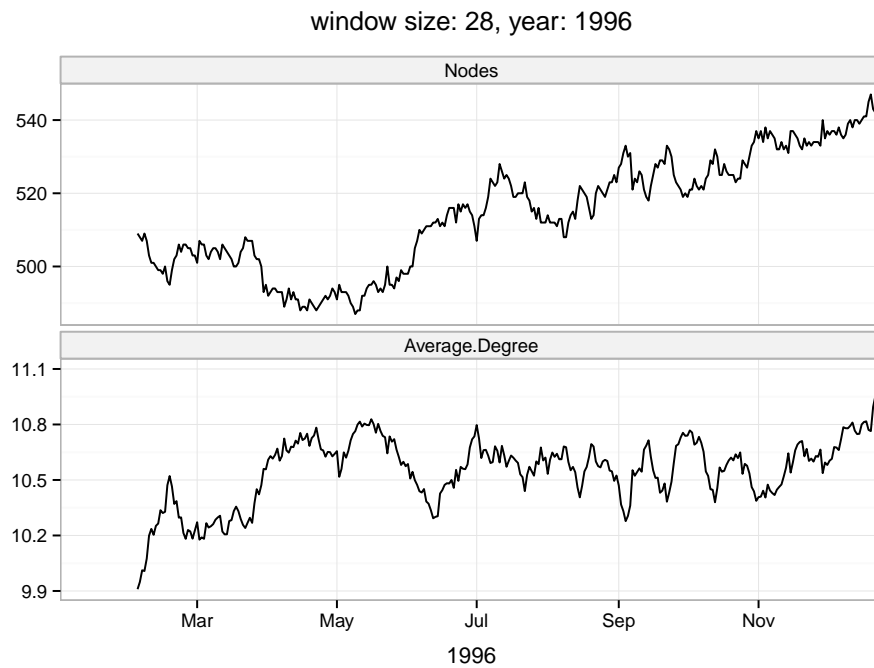


Figure 8: Compared evolution of the number of nodes of the TG-Graph and of its average nodes degree for 1996.

window size: 28, year: 2006

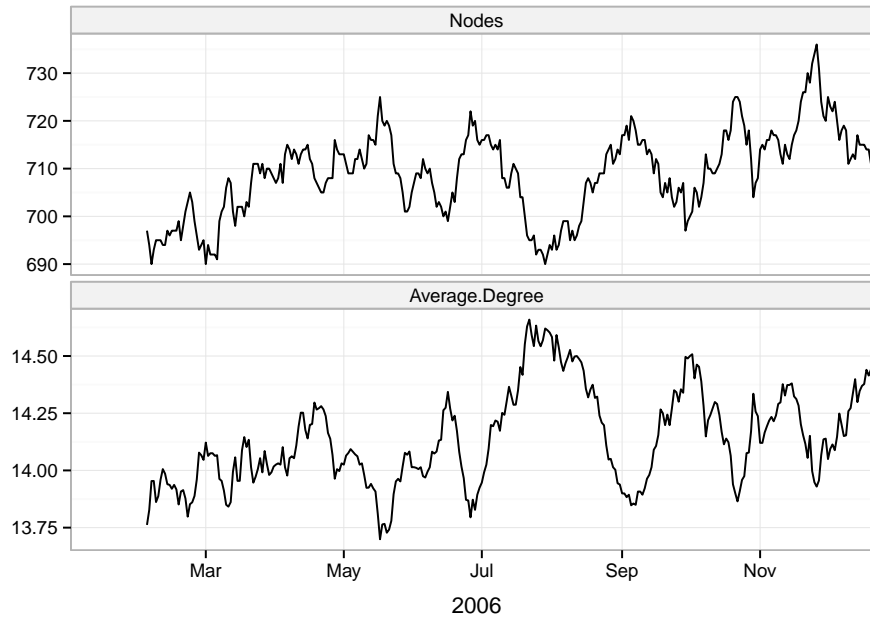


Figure 9: Compared evolution of the number of nodes of the TG-Graph and of its average nodes degree for 2006.

Another set of measures seems to be linked, the number of nodes and the average path length. When studying the graphics it appears that, the faster the increase of the number of nodes, the slower the decrease of the average path length as illustrated on Figure 10.

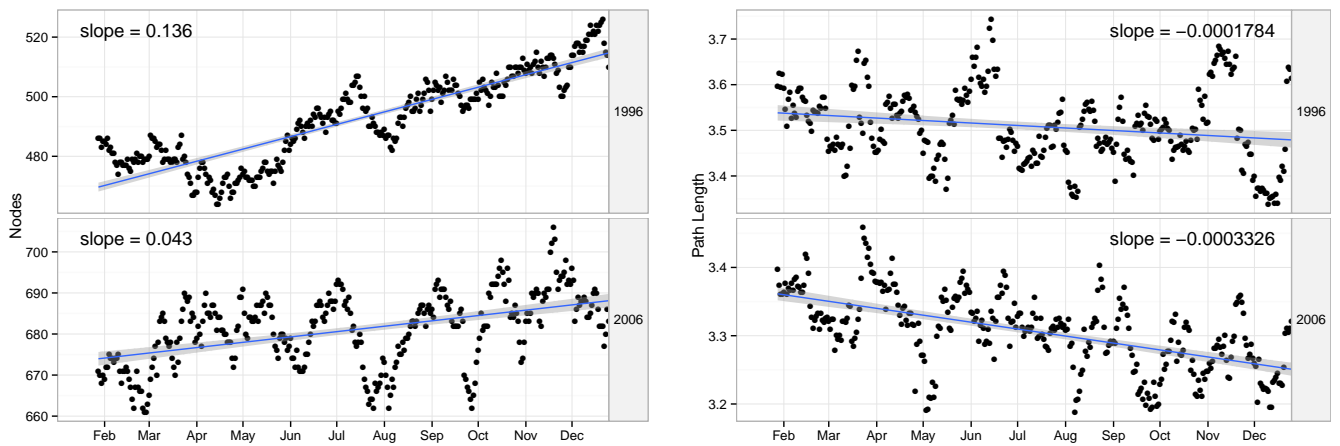


Figure 10: The faster the increase of the number of nodes, the slower the decrease of the average path length (for both years).

From a graph perspective, when new nodes are added to a graph, and if they are sparsely linked to other nodes (low degrees), these arrivals entail an increase of the average path length. Conversely, if new links are added to the graph, some paths between nodes are shorter which decreases the average path length. We thus have two opposite processes participating in the network growth. The comparison between 1996 and 2006 graphics suggests that the proportion of new nodes compared to new links was more important for the shipping network of 1996 than it was for 2006. In other words this means that in 1996 the main force driving the growth of the shipping network was the expansion while in 2006 it was the densification. We should however warn the reader because the slope values on which this interpretation is built are very small, but the tendencies were the same for every time

granularity values.

4 Temporal Networks

The previous sections focused on the analysis of the maritime dataset when viewed as static graphs or sets of static graphs, with nodes being the ports and arcs indicating a possible route between two ports. Within such a static network time is not present. Useful timed information could feed the network analysis. Indeed, from the Lloyd’s dataset, an arc could actually hold many information: number and dates of departures on that arc, their distribution through time, arrival dates, as well as measures related to the duration of the journeys on this arc (distribution, average, median, *etc.*). Classical complex networks usually don’t contain time related information. In this section we introduce temporal networks, a graph model that contains all time-related events of ships’ movements.

4.1 The temporal network model

The graphs considered in the two previous sections are 1-graphs, meaning that whatever the number of ships moving from one port A to the next call C , there will be only one arc linking A to C . If we want to represent all the ships moving from A to C we should model it using a multigraph. In [28] the authors consider a multigraph based on the notion of route, ”characterized by a predetermined sequence of ports, arrival schedule, frequency and deployed ships”, and the combination of such routes forms the directed maritime networks, but they restrict their analysis to the top 20 shipping lines for the period 1995-2011. Olivier Joly also raised this question of multigraph modeling but as he wrote, his study took place in a long-term perspective such that, adding one arc per ship, would have been an unnecessary complication of the problem. Remains the question of time. As he noticed, such a graph makes the assumption of synchronicity of ships’ movements and of the fact that ships are moving together and simultaneously [16].

For tackling this problem we propose to consider a graph model able to carry time information: the temporal networks (or evolving graphs). From the data, ships movements, we can build such a graph by associating one node to each port and one arc between two ports A and C as soon as at least one ship, during the considered period of time, moved from port A to the next one C . If the date of the departure from A was t , then the arc is valuated with t . More generally, if there are n dates of departures of ships from port A to C during the period, then the arc (A, C) will be valuated by the n dates.

The TG-Graphs presented in previous section can be considered as special cases of temporal networks. Indeed, a $TGG(t, J)$ is a non valuated sub-graph of a temporal network T , a sub-graph that only contains the set of arcs of T that are labelled by dates $d \in [t, t + J - 1]$, and, of course, corresponding nodes.

Numerous models exist in the literature for time enable networks. These models are presented under various names but the term *Temporal Network* tends to be the most accepted [13]. The general idea behind temporal networks is that they have a classical network structure, except that each edge (or arc) is *labelled* with a list of dates (or intervals) that indicate *when* that edge is *active*. The labelling of edges is commonly used in distributed computing and is close to the idea of a weighted graph where a value qualifies edges. When an edge (or arc) is active it behaves like a classical edge. When it is not active it should be considered non-existing. Consider a node A connected to a node B , and B connected to a node C . In a classical network, A and C are indirectly connected through B . In a temporal network time constraints apply on edges. It means that the link (A, B) needs to be active first and then link (B, C) in order to consider that A and C are indirectly connected. In this case even if edges are undirected, a path existing from A to C does not prevails a path from C to A . Edges (A, B) and (B, C) must be *active*, for the path to exist. A path in a network with such

time constraint is called a *journey* [33], a *non-decreasing path* or a *time-respecting path* [19].

Figure 11 illustrates a temporal network. A *journey* exists in this network from node A to node E because time constraints are respected. Indeed the path $\{A, B, C, D, E\}$ can be traversed with respect to time (for instance, (A, B) at time 1, then (B, C) at time 1, then (C, D) at time 2, and finally (D, E) at time 4). The exact opposite path is not a valid journey. Finally the whole concept of *path* and subsequent measures (diameter, radius, eccentricity, and most centralities) are redefined by temporal constraints.

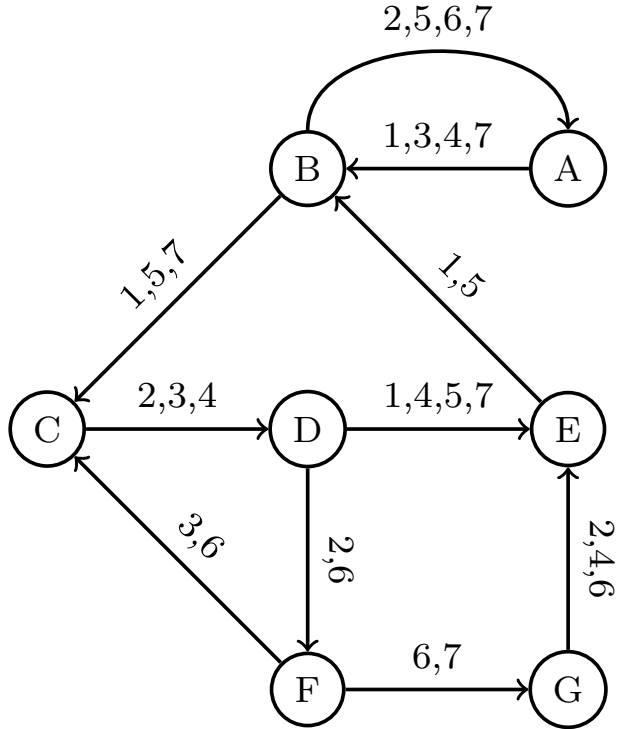


Figure 11: An example of *temporal network* with labelled edges for time steps when they are active.

Note that in this example no stop on nodes is assumed, travelling times of edges are neglected, and only the *activity* of the edges is considered. In the sequel, both stops on nodes and travelling times of edges will be taken into account for the computation of the journeys.

4.2 Least Cost Journeys in Maritime Temporal Networks

We propose here to adopt a new point of view for studying maritime networks. Indeed, temporal information and related graph algorithms allow the formulation of new problems. Instead of analyzing shipping networks from the port or the vessels point of views, we can look at the network from the container point of view. Let us consider that a given container located at a given place needs to reach a destination. To that purpose it will be shipped through one or several vessels in order to reach the destination. Under various assumptions multiple constraints can be considered. From Lloyd's dataset it is assumed that each arc of the network is *labelled* with the following information:

- the list of departure dates with each vessels' departure on this arc (from source port to destination port);
- the list of travel duration associated with the mentioned departures on this arc;
- the average, standard deviation, and median travel duration on this arc.

Based on these elements it is possible to consider new problems for the shipment of one container from one place to another.

1. Minimisation of transshipments. The number of different calls and subsequent need for transshipment has an inevitable economic cost that could be considered for minimisation, such that a journey with a minimum number of transshipments could be more desirable, at the cost of an overall longer journey.
2. Minimisation of the arrival date. If the wish it to ship the container as fast as possible, then the overall duration from the moment we try to ship the cargo has to be minimised.
3. Minimisation of the time at sea. It corresponds to finding a fast journey but when the start date does not matter. This case is useful for instance for renewable resources with a freshness to be maximized.

Those three opened questions remain to be investigated using the temporal network framework in the field for maritime networks analysis. The remainder of the section focuses on the minimization of the arrival date.

4.3 Horizon and Foremost Journeys

Various adaptation of the classical shortest path problem have been widely investigated in the field of temporal networks [33].

As stated, we consider here the list of departures on each arc as a constraint. For sake of statistical relevance in regards to some extreme (and maybe erroneous) values with wide standard deviation, we chose to consider the travel duration on an edge as the median value of the observed real durations. Moreover the minimum stay at a location for a container is set to one day in order to account for transshipment delay.

In the case of maritime network, before considering the optimisation of the arrival date we first consider if an arrival is ever possible during this observation window. Since the datasets are available for periods of one full year, then one year is the maximum accepted duration for a journey. We define for each port, seen as a potential starting point for a container, the number of destinations that it can reach in average throughout the duration of the observation (one year). Being able to reach another port means that a journey (a path that respects time constraints) leading to this port must exist.

We call *horizon* the average number of reachable destinations for a given port. When time is not considered, the horizon of a port is expected to be large, probably on the order of the size of strongly connected component this port is in. When time is considered, due to time constraints, this number is expected to be less. Some ports have been identified in the previous section to be leaves in the giant connected components. They are reachable but cannot reach any destinations. Those 32 ports in 1996 (respectively 34 in year 2006) have no horizon and a potentially infinite journey duration to any destination.

Moreover the duration of journeys can also be investigated. Time and duration are usually considered (mixed-up) with weights such that a shortest path on a classical network with arcs durations used as weights is computed. This gives an artificial value for the journey duration since actual vessels departures are omitted. We compute here the actual Foremost journeys in the network and compare them with this naïve approach.

Figure 12 shows an important difference in average for the horizon when time is truly considered for reaching destinations. On average, in 1996, the horizon is 820 nodes, which is close to the size of the giant strong component (835), whereas it is on average 460 when time is considered with a wide standard deviation (± 130). A similar difference is noted on the 2006 dataset.

The difference between the temporal and atemporal approaches for average journey duration is very important. While the naïve *weighted* shortest path gives an average journey duration of 20 ± 47 days (resp. 17 ± 42 for year 2006), the time respecting foremost journey computation gives an average of 112 ± 55 (resp. 96 ± 54).

That difference between the supposed classical time consideration on the static network and the effective computation of time dependent paths is of paramount importance for the analysis of

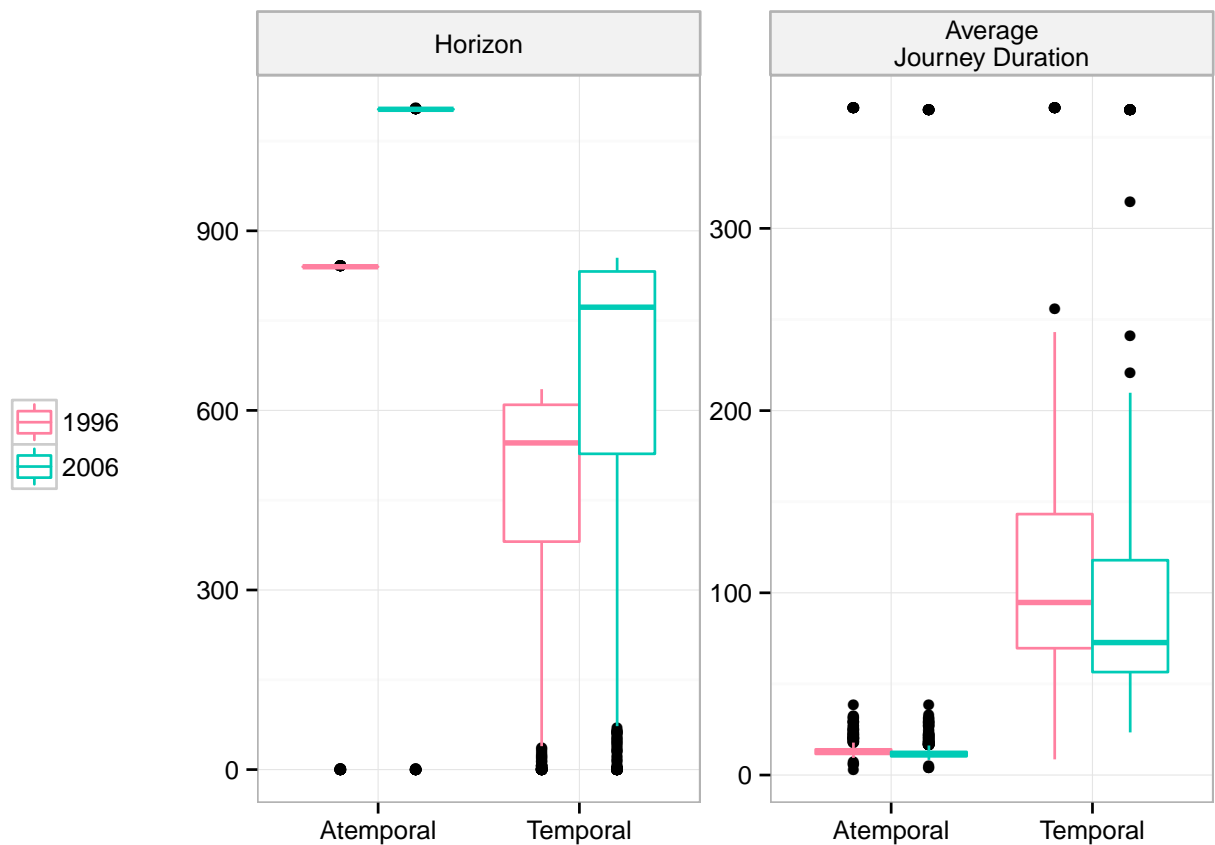


Figure 12: Temporal versus Atemporal Analysis of average journey durations and ports horizon. When atemporal, the horizon (in number of nodes/ports) is in the order of the size of the connected component of each node and the average journey duration (in days) only considers arcs median duration as a simple weight. When temporal, the horizon is inferior to the Atemporal and the average journey duration is way more important than the naïve *weighted* shortest path. Years 1996 and 2006 are investigated and show similar behaviour.

maritime networks. While it is commonly believed that the maritime network is strongly connected and that goods can cross the globe in about 20 days, this analysis shows a different point of view. If indeed the network shows Scale-Free properties, ports that are not directly connected to the large hubs are actually isolated.

5 Conclusion

Classical analyses of maritime networks neglect the temporal information contained within the data. We put the focus on the introduction of these temporal information within the graph models that can be built from ships' movements recorded in data sources like the Lloyd's Lists Periodicals.

We have considered three different graph models, starting from the classical static one that gather all ships' movements into one unique graph. It has been shown that the network carries expected properties like small-worldness and scale-freeness (to an extent).

The second model called TG-Graph, stands for Time Granularity-based Graph, is characterized by a time granularity allowing to build time series of TG-Graphs for the study of the evolution of many metrics during a given period of time. We have shown that the temporal evolution of simple metrics like graph order or the compared evolution of couples of metrics can bring some interesting insights for the understanding of the underlying shipping network.

The introduction of time granularity for building TG-graphs raises multiple questions, about the choice of the granularity, about the existence of a best granularity value for studying some phenomenon, about the sensitivity of metrics to granularity changes or about the consistence of results obtained from different granularity values. At this stage of the study we are not able to answer these questions and in this chapter we've just opened the discussion about the relevance of such a graph model. We focused on a few different time granularities (1, 7, 14, 21 and 28 days) and we have restricted our study to global metrics. However, similar studies can be driven for nodes and edges. These analyses could lead to the comparison of the evolution, the stability, and the robustness of a chosen subset of ports at different periods of time or for observing their response to specific events like crises for instance.

Temporal network was our last model. It takes into account time more explicitly by including time related data directly in the graph. This model allows a new point of view in the analysis of the maritime networks. We can now formulate problems and answer questions related to how cargo can be shipped through the maritime network, what path to use and when to use them in order to optimise the transit of goods for different optimization functions. The analysis of these graphs brought new results that change the way those networks should be considered. Despite its small-world property and its hubs, when time is considered, the network shows average journey duration that are 5 times longer than what a standard *off-time* network would have guessed.

6 Acknowledgements

The authors would like to thank César Ducruet and his team for providing us with their data and for their availability for the questions we had about these data.

References

- [1] S. Alumur and B. Y. Kara. Network hub location problems: The state of the art. *European Journal of Operational Research*, 190(1):1 – 21, 2008.
- [2] A. Clauset, C. Shalizi, and M. Newman. Power-Law Distributions in Empirical Data. *SIAM Review*, 51(4):661–703, Nov. 2009.
- [3] C. Ducruet and T. Notteboom. The worldwide maritime network of container shipping: spatial structure and regional dynamics. *Global Networks*, 12(3):395–423, 2012.

- [4] C. Ducruet, C. Rozenblat, and F. Zaidi. Ports in multi-level maritime networks: evidence from the atlantic (1996–2006). *Journal of Transport Geography*, 18(4):508 – 518, 2010. Special Issue on Comparative North American and European gateway logistics.
- [5] C. Ducruet and F. Zaidi. Maritime constellations: a complex network approach to shipping and ports. *Maritime Policy & Management*, 39(2):151–168, 2012.
- [6] A. Dutot, F. Guinand, D. Olivier, and Y. Pigné. Graphstream: A tool for bridging the gap between complex systems and dynamic graphs. In A. Alaoui and C. Bertelle, editors, *Proceedings of Emergent Properties in Natural and Artificial Complex Systems. Satellite Conference within the 4th European Conference on Complex Systems (ECCS'2007), October 4-5. Dresden, Germany.*, pages 63–72, 2007.
- [7] P. Erdős. On Random Graphs I. *Publicationes Mathematicae (Debrecen)*, 6:290–297, 1959.
- [8] A. Ferreira. On models and algorithms for dynamic communication networks: The case for evolving graphs. In *4^e rencontres francophones sur les Aspects Algorithmiques des Telecommunications (ALGOTEL'2002)*, Mèze, France, 2002.
- [9] A. Frémont. Global maritime networks: the case of maersk. *Journal of Transport Geography*, 15:431–442, 2007.
- [10] W. L. Garrison and D. F. Marble. Factor-analytic study of the connectivity of a transportation network. *Papers in Regional Science*, 12:231–238, 1964.
- [11] S. Gelareh, S. Nickel, and D. Pisinger. Liner shipping hub network design in a competitive environment. *Transportation Research Part E: Logistics and Transportation Review*, 46(6):991 – 1004, 2010.
- [12] C. S. Gillespie. Fitting Heavy Tailed Distributions: The poweRlaw Package. *Journal of Statistical Software*, 64(2):1–16, 2015.
- [13] P. Holme and J. Saramäki. Temporal networks. *Physics Reports*, 519(3):97–125, Oct. 2012.
- [14] Y. Hu and D. Zhu. Empirical analysis of the worldwide maritime transportation network. *Physica A: Statistical Mechanics and its Applications*, 388(10):2061 – 2071, 2009.
- [15] A. Imai, K. Shintani, and S. Papadimitriou. Multi-port vs. hub-and-spoke port calls by container ships. *Transportation Research Part E: Logistics and Transportation Review*, 45(5):740 – 757, 2009.
- [16] O. Joly. *La structuration des réseaux de circulation maritime : position des plates-formes d'interconnexion en Europe du Nord-Ouest*. PhD thesis, Le Havre University, 1999.
- [17] P. Kaluza, A. Kölzsch, M. T. Gastner, and B. Blasius. The complex network of global cargo ship movements. *Journal of the Royal Society Interface*, 2010.
- [18] K. J. Kansky. *Structure of transportation networks : relationships between network geometry and regional characteristics*. PhD thesis, University of Chicago, 1963.
- [19] D. Kempe, J. Kleinberg, and A. Kumar. Connectivity and Inference Problems for Temporal Networks. In *Proceedings of the Thirty-second Annual ACM Symposium on Theory of Computing, STOC '00*, pages 504–513, New York, NY, USA, 2000. ACM.
- [20] F. G. Laxe, M. J. F. Seoane, and C. P. Montes. Maritime degree, centrality and vulnerability: port hierarchies and emerging areas in containerized transport (2008–2010). *Journal of Transport Geography*, 24(0):33 – 44, 2012. Special Section on Theoretical Perspectives on Climate Change Mitigation in Transport.

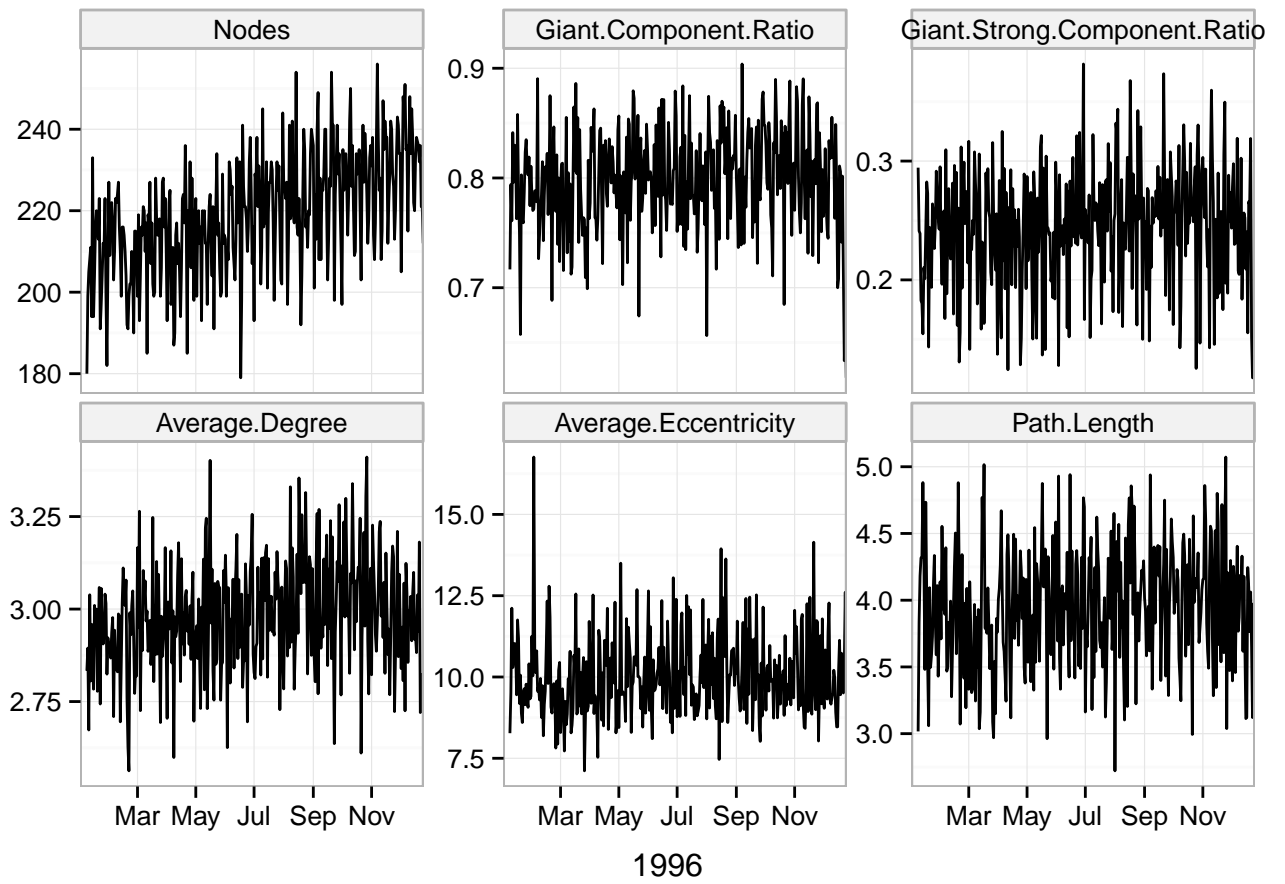
- [21] M. Newman. Power laws, Pareto distributions and Zipf’s law. *Contemporary Physics*, 46(5):323–351, Sept. 2005.
- [22] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.
- [23] M. E. O’Kelly and H. J. Miller. The hub network design problem: A review and synthesis. *Journal of Transport Geography*, 2(1):31 – 40, 1994.
- [24] M. J. F. Seoane, F. G. Laxe, and C. P. Montes. Foreland determination for containership and general cargo ports in europe (2007–2011). *Journal of Transport Geography*, 30(0):56 – 67, 2013.
- [25] J. Sienkiewicz and J. A. Hołyst. Statistical analysis of 22 public transport networks in poland. *Phys. Rev. E*, 72:046127, Oct 2005.
- [26] Q. K. Telesford, K. E. Joyce, S. Hayasaka, J. H. Burdette, and P. J. Laurienti. The Ubiquity of Small-World Networks. *Brain Connectivity*, 1(5):367–375, Dec. 2011.
- [27] N. Tran and H.-D. Haasis. Literature survey of network optimization in container liner shipping. *Flexible Services and Manufacturing Journal*, pages 1–41, 2013.
- [28] N. Tran and H.-D. Haasis. Empirical analysis of the container liner shipping network on the east-west corridor (1995–2011). *NETNOMICS: Economic Research and Electronic Networking*, 15(3):121–153, 2014.
- [29] Q. H. Vuong. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica*, 57(2):307–333, Mar. 1989.
- [30] C. Wang and J. Wang. Spatial pattern of the global shipping network and its hub-and-spoke system. *Research in Transportation Economics*, 32(1):54 – 63, 2011. Container Transport Operations.
- [31] D. J. Watts and S. H. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393(6684):440–442, June 1998.
- [32] D. Wei-Bing, G. Long, L. Wei, and C. Xu. Worldwide marine transportation network: Efficiency and container throughput. *Chinese Physics Letters*, 26(11):118901, 2009.
- [33] B. Xuan, A. Ferreira, and A. Jarry. Evolving graphs and least cost journeys in dynamic networks. In *WiOpt’03: Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*, page 10 pages, 2003.

7 Appendix

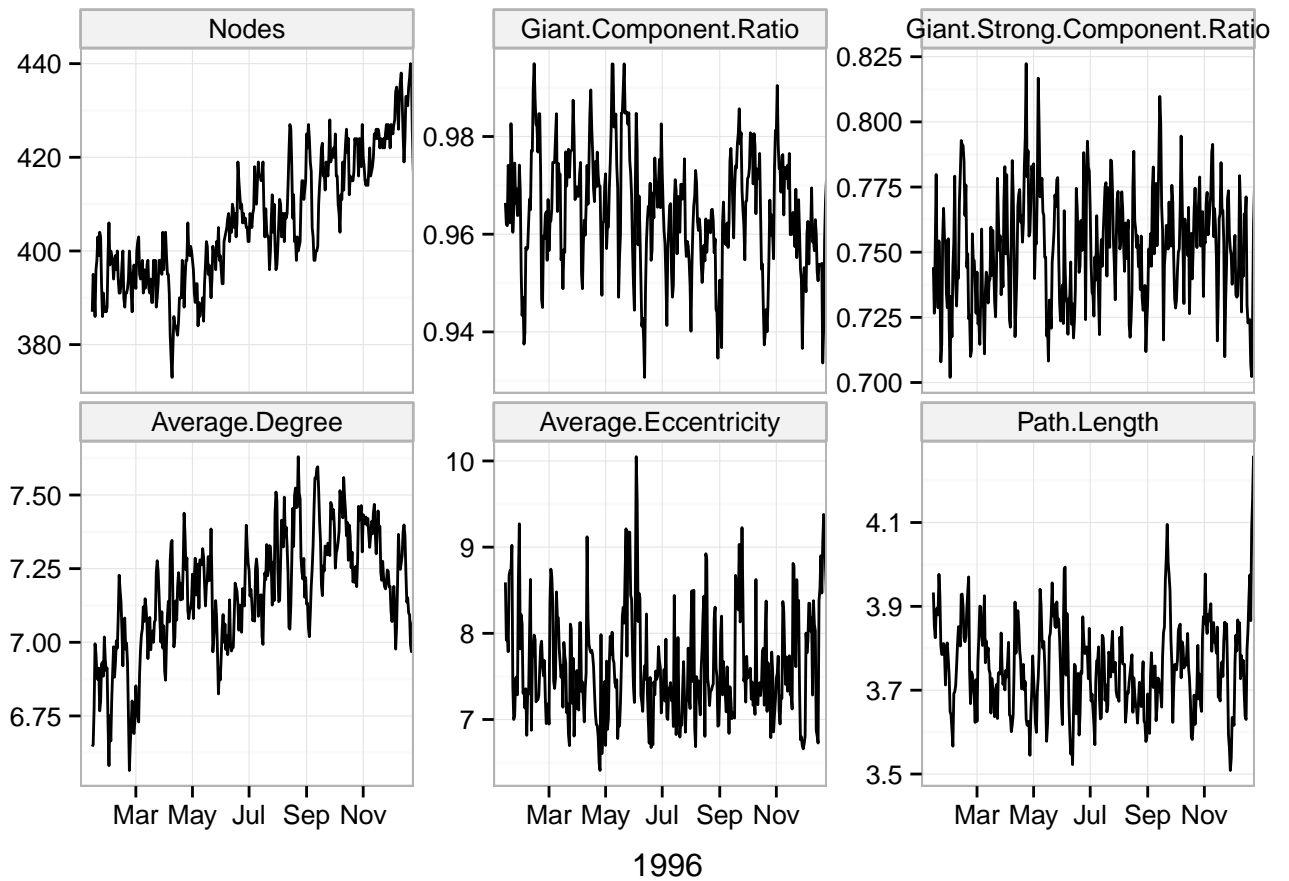
All experiments were conducted using GraphStream [6], a tool for designing, analysing and visualising dynamic graphs.

The following figures illustrate graph measures with various granularity sizes for the two datasets.

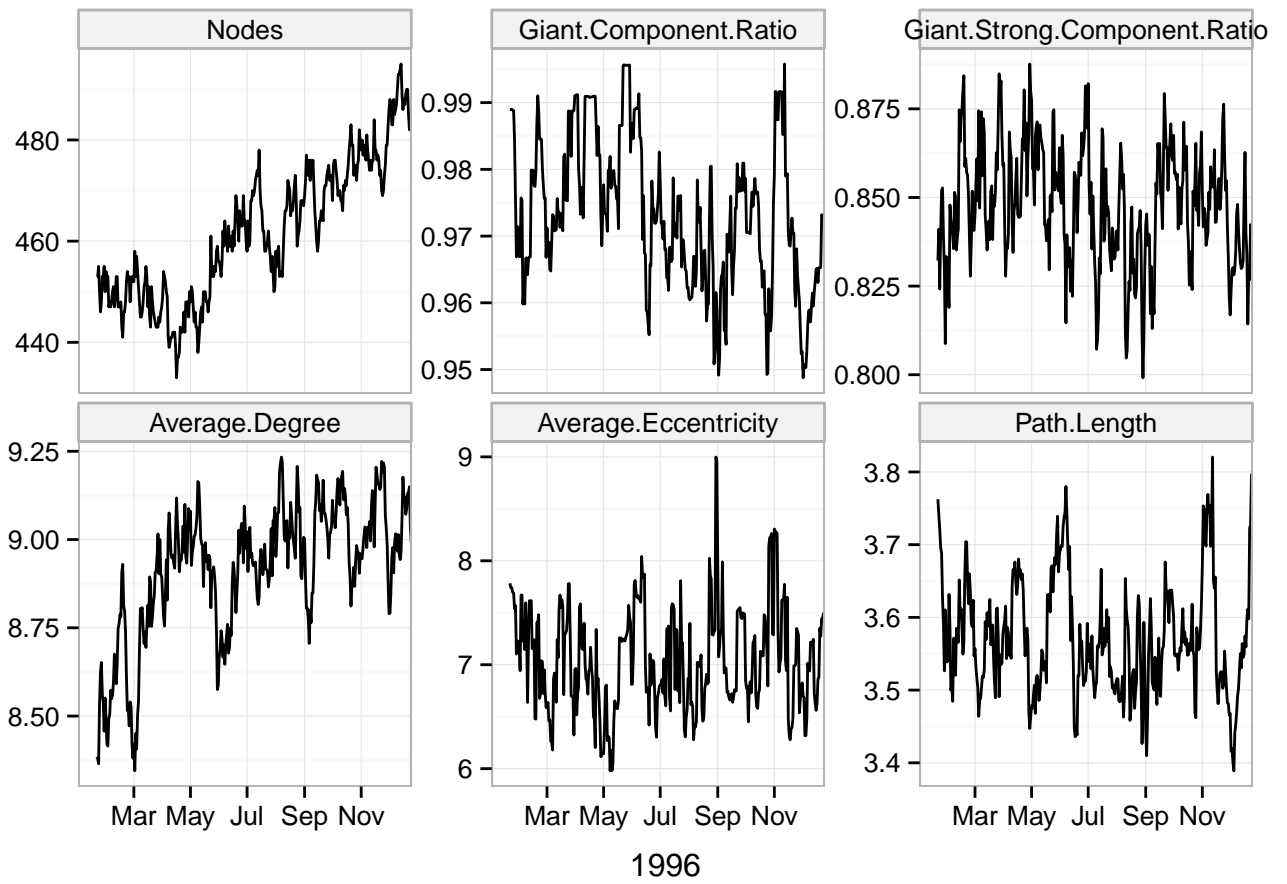
window size: 1, year: 1996



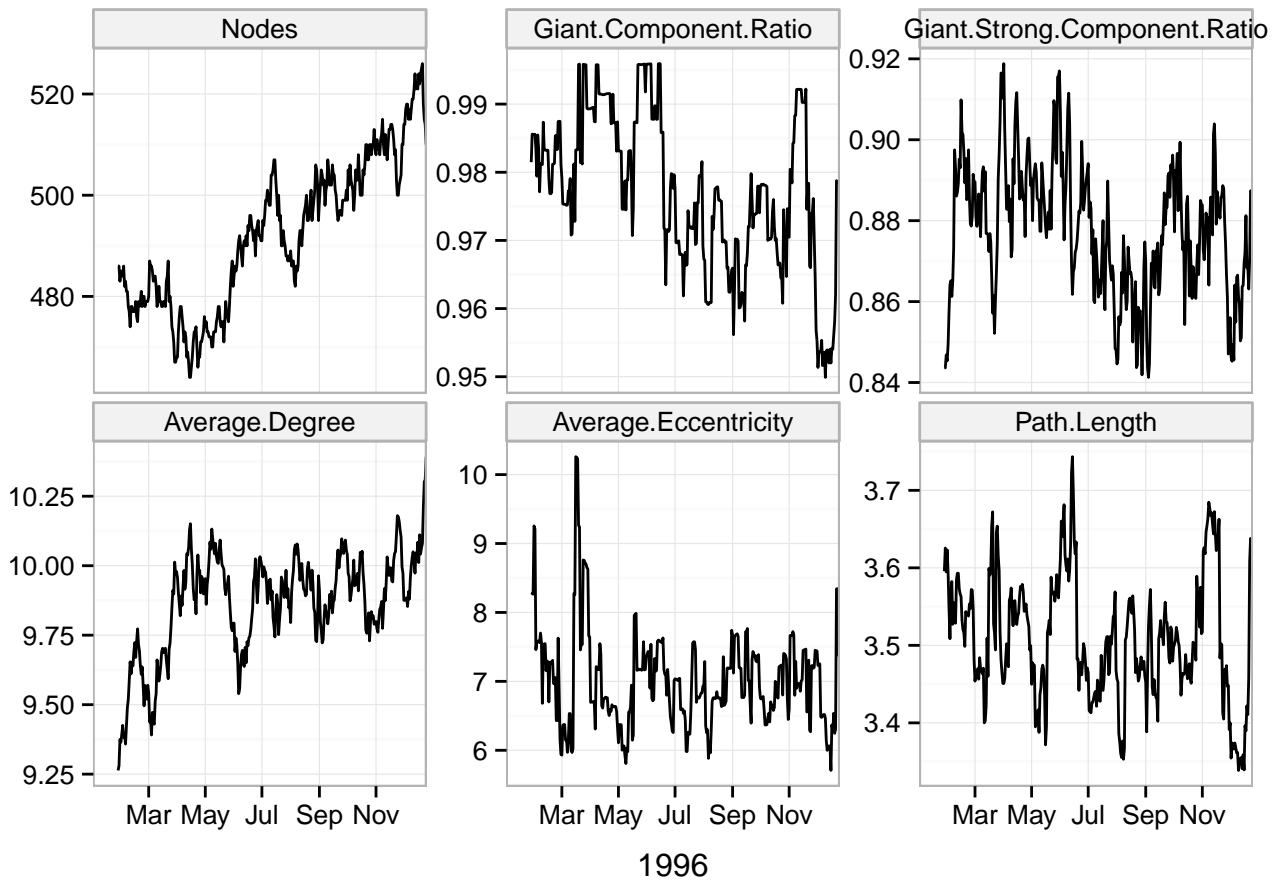
window size: 7, year: 1996



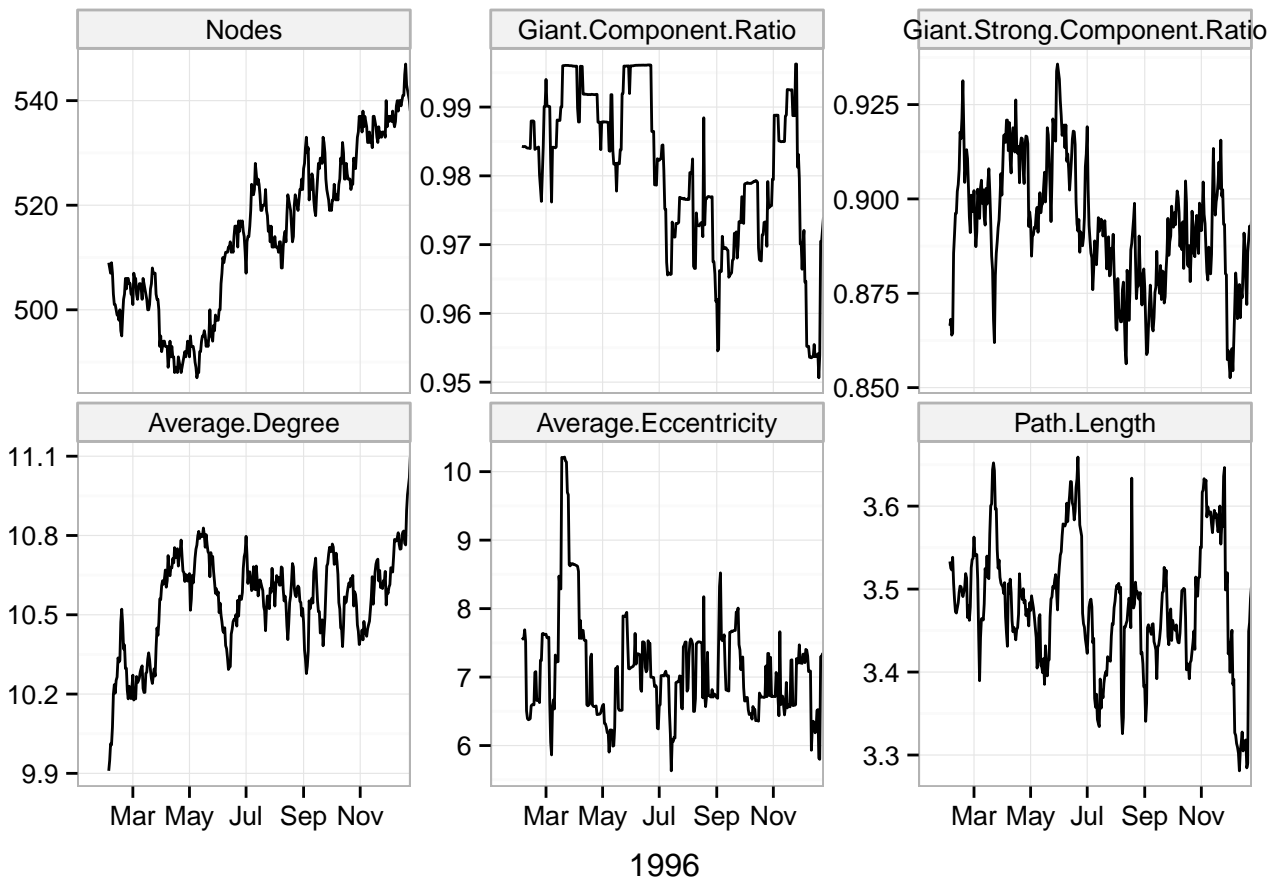
window size: 14, year: 1996



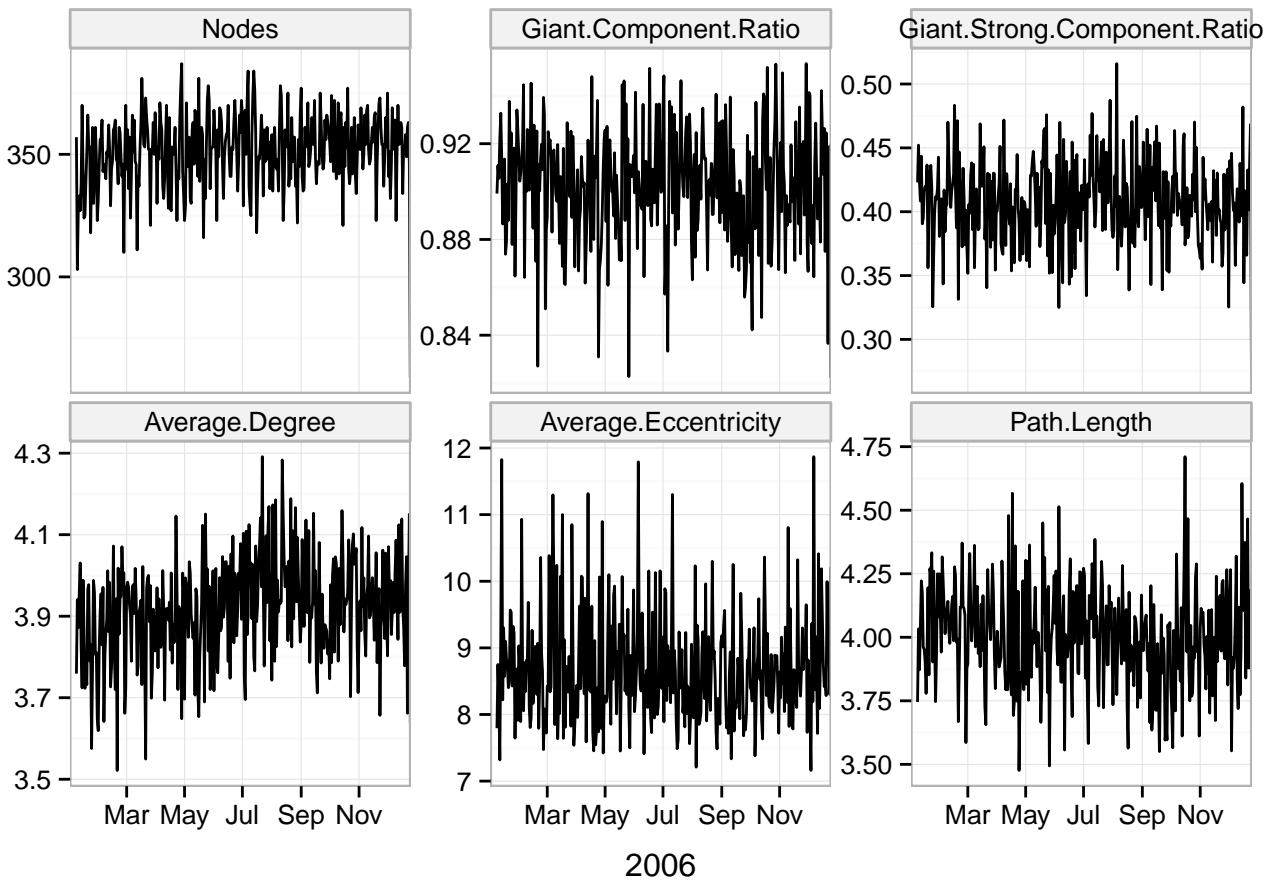
window size: 21, year: 1996



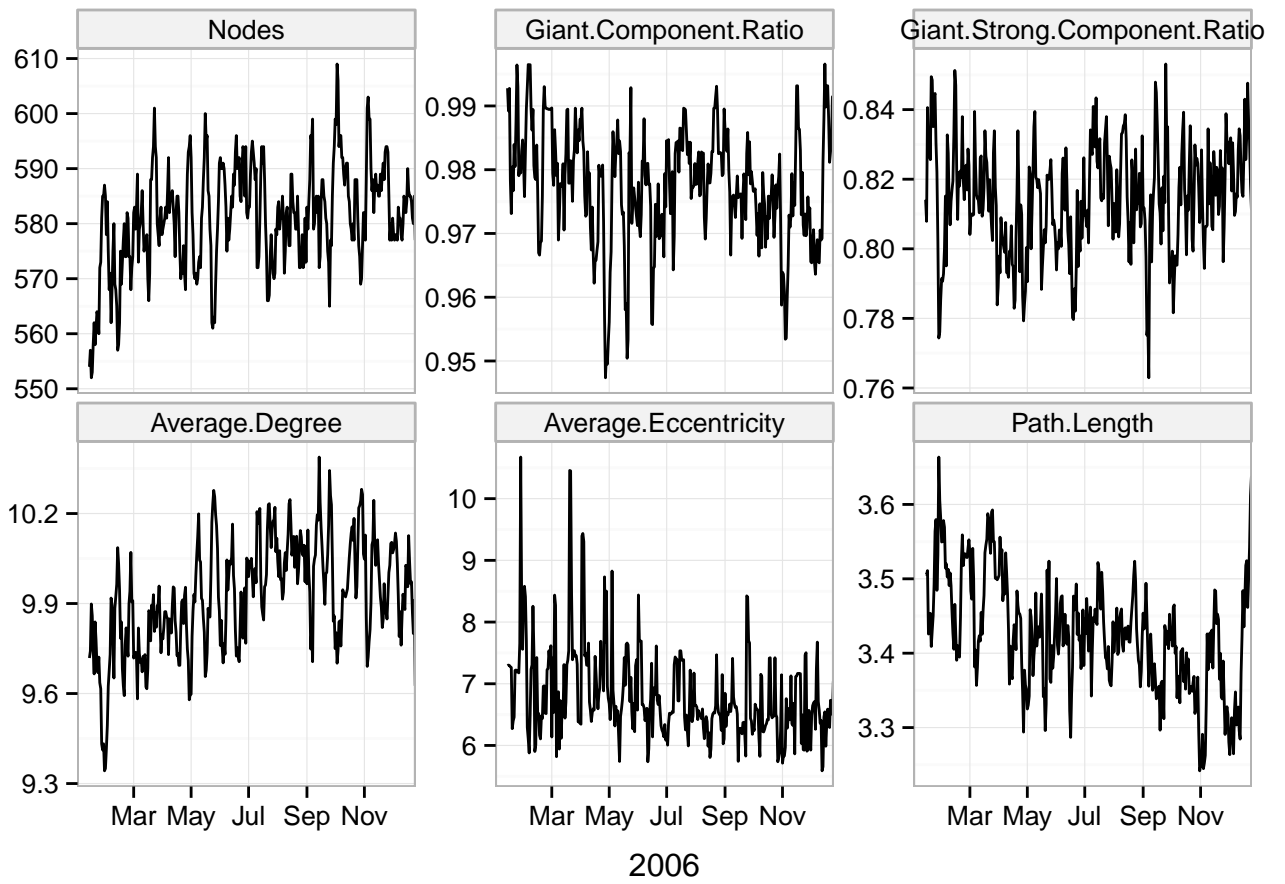
window size: 28, year: 1996



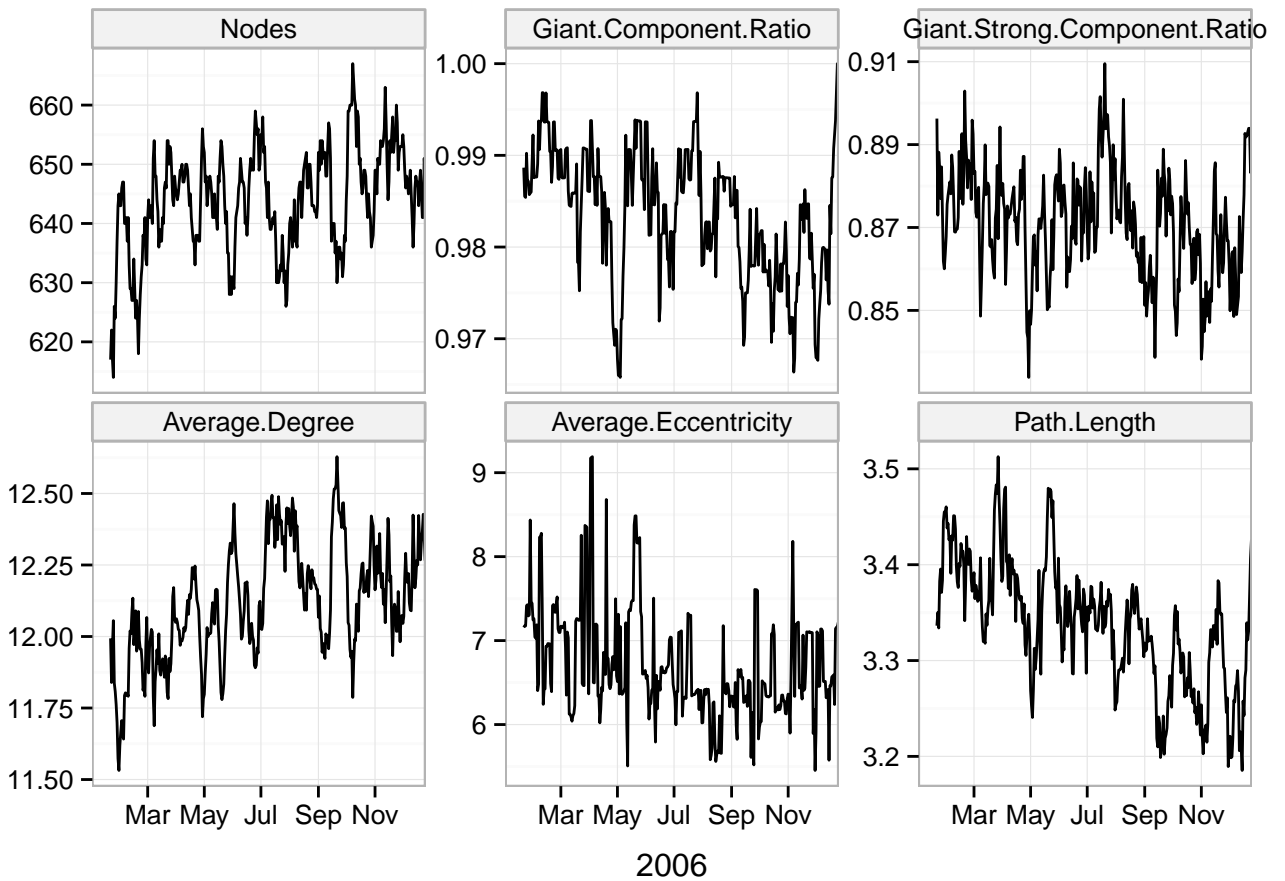
window size: 1, year: 2006



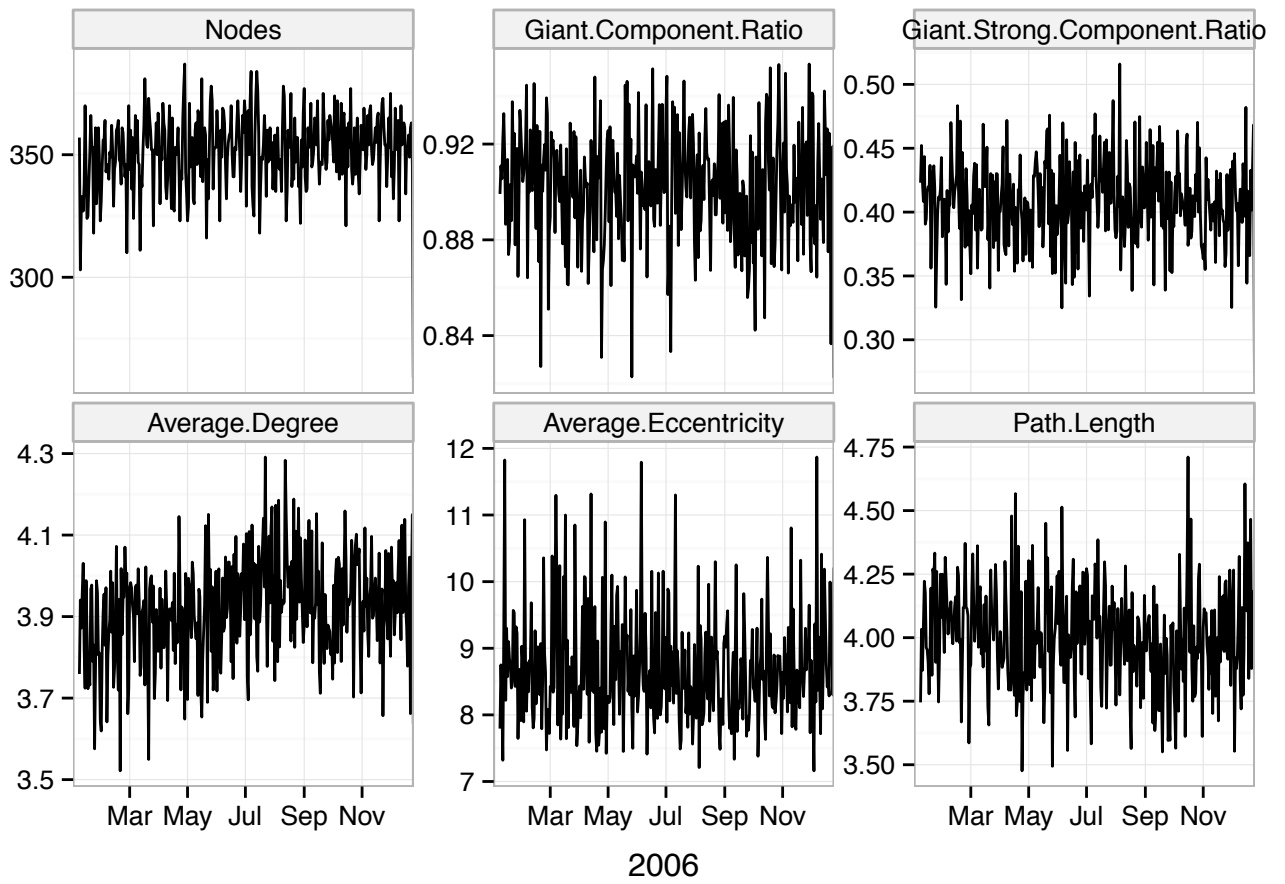
window size: 7, year: 2006



window size: 14, year: 2006



window size: 1, year: 2006



window size: 28, year: 2006

