



HAL
open science

Comparing Static and Dynamic Graphs built from Mobility Traces

François Queyroi

► **To cite this version:**

François Queyroi. Comparing Static and Dynamic Graphs built from Mobility Traces. MARAMI 2019, Nov 2019, Dijon, France. hal-02181680

HAL Id: hal-02181680

<https://hal.science/hal-02181680v1>

Submitted on 12 Jul 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.

Comparing Static and Dynamic Graphs built from Mobility Traces

A case study with maritime transportation data

François Queyroi

the date of receipt and acceptance should be inserted later

Keywords sequential data analysis ; dynamic graphs ; time-dependent shortest-paths; transportation networks

Introduction. This paper introduces ongoing researches on graphs built from aggregation of sequential data. More precisely, we focus on spatially and temporally located sequential data such as the records of ports visited by transportation vessels. The traces used here are extracted from the Lloyd's List Intelligence database¹ and consists of container ships movements between April 1st 2009 and July 31th 2009 (made available by the WorldSeastems ERC project [3,4]): the ports visited along with the arrival and sailing dates of the vessels.

Although dynamic in nature, the shipping network is often studied as a static graph built by aggregation over a given period [5,7]. For instance, [7] compared two static aggregations of transportation companies lines in 2004 and 2014. Looking at the robustness of some graph measures, they concluded that the maritime network became more vulnerable to port failures.

However, there are several ways to construct a network from sequential data. They actually correspond to different aggregation strategies which lead to different interpretations of a same graph measure. The work presented here focuses on the comparison between network statistics depending on the chosen strategies. One important contribution of this work is the comparison between static shortest-path lengths and time-dependent fastest-path durations.

Aggregation strategies. In the context of maritime transportation networks, we use the concepts of *space L* and *space A* (using a terminology close to [5]). For a vessel with a sequence of visited ports ($a \rightarrow b \rightarrow c$), a *space L* network contains direct links ($a \rightarrow b$) and ($b \rightarrow c$) while a *space A* network also contains the indirect link

François Queyroi
LS2N, UMR CNRS 6004, Université de Nantes
E-mail: francois.queyroi@univ-nantes.

¹ Data and algorithms' implementations used are available at https://github.com/fqueyroi/tulip_plugins/tree/master/papers/NetFromSequentialData

($a \rightarrow c$). The out-degree of a vertex (port) a in *space L* (respectively *space A*) corresponds to the number of direct destinations (resp. reachable ports) using a ship leaving a . The unweighted shortest-path distance between ports a and b in *space L* (resp. *space A*) corresponds to the minimum of number of visited ports (resp. number of ships used) from a to b .

We can also use knowledge about the domain to define networks in *space P*. Indeed, vessels are not independent agents but are assigned to a set of ports (*lines* or routes) by shipping companies. The assignments may change over time. Other researches [7, 5] directly use the lines given by shipping companies to define *space P*. For a given line, *space P* includes the indirect links between all ports visited on this route. Unlike *space A*, it does not include indirect links between ports that are never part of the same line and where no exchange of goods could have taken place. In *space P*, the out-degree corresponds to the number of lines available in a port while distances correspond to the minimum number of different lines required to go from a port to another.

However, the Lloyd's database does not include shipping companies' planned lines but "only" the temporal sequences of ports actually visited by each vessel. The sequences still contain some patterns that can be used to recover the lines. For instance, a classic shipping route passes through a given port at most twice [6] (following a circle pattern). We develop a simple procedure to extract lines from the sequence of ports visited by a given ship: we recursively cut the port sequence using as pivot the most visited port until the sequence is partitioned into valid circle sub-sequences.

Notice that temporal and sequential information are lost using *space L*, *P* or *A*. However, we can use ships' arrival and sailing dates in each port and set the average time needed to cross an edge ($a \rightarrow b$) as the average difference between sailing dates from a and arrivals dates in b . But this definition of edges weight still discards an important feature of temporal networks namely the frequencies of interaction between elements (number of ships going between two ports on a given period).

Another aggregation that preserves frequencies is the use of dynamic graphs model. The idea is to store every arrival and sailing dates for each edge in *space L, P* or *A*. The notion of shortest-path distance can be intuitively interpreted the fastest-path duration which can be computed using a variation of Dijkstra's algorithm [1]. Notice that a path is here defined not only by a starting port but also by a starting date.

This aggregation *a priori* also discards sequential information as one path can jump from one ship to another without any cost. This hidden hypothesis leads to strange results as a fastest path may include a lot of ports. We can introduce a stopover cost using multi-modal fastest path durations [2]. One solution here is computing fastest-paths using sailing and arrivals dates on edges in *space P* or *A* adding a fixed time c (time required to take another line or ship) each time a port is visited. Indeed, in *space P* (resp. *A*), if a fastest path contains an internal node it means that a change of line (resp. ship) had to take place. The additivity of shortest-paths is therefore not violated and Dijkstra's algorithm variation can still be used.

Results We compare the four networks and computed distances using as vertex statistic the average distances from one port to the others. A low value indicates that a port is easily reachable. As the dynamic network may not be “temporally connected” we filtered out ports with a low frequency of interactions and took the average of travel duration over the period. We therefore compare the results on 438 ports. The density of connections for *space L,P* and *A* are 0.03, 0.1 and 0.15 respectively.

First results suggest that *space P* and *space A* are actually very similar w.r.t. static distances and almost identical w.r.t. temporal distances. This may indicate that most ships follow regular routes on this relatively short period.

They also suggest that unweighted distances in *space L* (*i.e.* minimum number of visited ports) is a better approximation for travel durations than the unweighted distances in *space P* or *A* (minimum number of used lines or ships). Although correlations with *space P* and *A* distances exist, it indicates that assessing ports accessibility only using *space P* or *A* may discard important information. This effect is stronger when looking at average travel durations with a cost of stopover $c = 2$. These correlations seem weaker when looking at the weighted distances (*i.e.* using the average time needed to cross an edge). This observation is counter-intuitive and further tests are needed to validate it.

Ongoing work and future directions. More work is needed in order to assess the effects of errors in the database. Indeed, some trips between distant ports are shorter than they should. This may affect weighted distances and fastest-paths computation. Moreover, our definition of *space P* relies on one possible route extraction algorithm that should be compared against others or directly against the shipping lines given by the shipping companies.

Future directions include the analysis of the measurements and their robustness to change on different time periods [7]. We will also compare other network measures (*e.g.* clustering coefficient) with their dynamic counterparts (*e.g.* a dynamic clustering coefficient corresponding to a local and dynamic version of betweenness centrality).

References

1. Dean, B.C.: Shortest paths in fifo time-dependent networks: Theory and algorithms. Report technique, Massachusetts Institute of Technology (2004)
2. Delling, D., Pajor, T., Wagner, D.: Accelerating multi-modal route planning by access-nodes. In: European Symposium on Algorithms, pp. 587–598. Springer (2009)
3. Ducruet, C.: ERC World Seastems project . <http://www.world-seastems.cnrs.fr/> (2014). [Online; accessed 17-June-2019]
4. Ducruet, C., Berli, J.: Mapping the globe. the patterns of mega-ships. Port Technology International **77** (2018)
5. Hu, Y., Zhu, D.: Empirical analysis of the worldwide maritime transportation network. Physica A: Statistical Mechanics and its Applications **388**(10), 2061–2071 (2009)
6. Shintani, K., Imai, A., Nishimura, E., Papadimitriou, S.: The container shipping network design problem with empty container repositioning. Transportation Research Part E: Logistics and Transportation Review **43**(1), 39–59 (2007)
7. Wang, N., Wu, N., Dong, L.L., Yan, H.k., Wu, D.: A study of the temporal robustness of the growing global container-shipping network. Scientific reports **6**, 34,217 (2016)