Journal of Transport Geography 69 (2018) 271–281

Contents lists available at ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrangeo

Transport network backbone extraction: A comparison of techniques

Liang Dai[a,b], Ben Derudder[b,*], Xingjian Liu[c]

[a] School of Public Administration, Nanjing University of Finance and Economics, China
[b] Department of Geography, Ghent University, Belgium
[c] Department of Urban Planning and Design, The University of, Hong Kong

ARTICLE INFO

Keywords:
Backbone extraction
One-mode networks
Complex network analysis
Southeast Asia

ABSTRACT

Network backbone extraction techniques reduce the size of networks while trying to preserve their key topological and spatial features. Various backbone extraction algorithms have been proposed in different scientific fields. Although of clear interest to transport geographers, backbone extraction techniques have been adopted unevenly and in an ad hoc fashion in transport geography research. In this paper we therefore present a conceptual and experimental comparison of backbone extraction techniques in a transport-geographical context, and explore the new insights each technique can offer to enhance our understanding of the Southeast Asian intercity air transport network (SAAN). We review six frequently-used methods, i.e. global weight thresholding method (GWTM), k-core decomposition method (KCDM), minimum spanning tree method (MSTM), primary linkage analysis method (PLAM), multiple linkage analysis method (MLAM), and the disparity filter algorithm method (DFAM), and elaborate their analytical essence by applying them to extract the backbone of the SAAN. The abstracted networks are compared in terms of their geographical and topological structures using the initial network as a benchmark. This comparison is then used to point out the different techniques' potential in light of different transport geography research applications.

1. Introduction

In recent decades, there has been burgeoning interest in the structural analysis of transport *networks* across modes and scales (e.g. Ducruet et al., 2010; O'Kelly, 2016; Wang et al., 2009; Liu et al., 2016). In these networks, nodes commonly represent spatial units such as cities, airports, ports, and stations, while edges identify transport-related interactions between the nodes. In addition, edges are typically weighted by capacity, frequency, distance, or the time it takes to "travel" between nodes. In theory, the application of the ever-expanding suite of network analysis techniques allows examining complex transport systems at the level of nodes and dyads as well as the network in its entirety (Barthélemy, 2011; Tsiotas and Polyzos, 2017). To date, network-focused research efforts in transport geography have primarily focused on four areas of enquiry: (1) the representation of non-planar and planar transportation systems through networks (e.g. Lin and Ban, 2013); (2) the analysis of the topographical and topological features of transport networks (e.g. Lin, 2012); (3) tracing the spatial and structural evolution/dynamics of these networks over time (e.g. Ducruet, 2017); (4) and modelling transport networks with the specific purpose of uncovering their underlying mechanisms (e.g. Zhang et al., 2016).

The visualization, description and analysis of transport networks continue to face a range of challenges. For example, the fact that transport networks are *spatial* networks where nodes are preferably visualized in their exact geographical location makes producing transport flow maps a complex proposition (Vertesi, 2008). Dense networks with locally/regionally clustered edges in particular pose challenges when trying to explicitly convey the overall structure (Hennemann, 2013). Furthermore, analytically trivial edges in a network may give rise to biases in the measurement and interpretation of network topologies (Radicchi et al., 2011). For these and a number of related reasons, it is often helpful to extract the "backbone" of a network: a simplified version that is reduced in size – i.e., some edges and/or nodes are deleted – but retains the most "valuable" information contained in the original network. The abstracted network can be mapped and explored with significantly less effort, and this without too much compromising the real-world remit of the network.

To achieve this goal, a large number of methods have been developed. These methods aim to de-densify networks by extracting their "backbone(s)", and range from simple thresholding (Derudder and Taylor, 2005) to more statistically-grounded methods such as disparity filter algorithms (Serrano et al., 2009). Needless to say, these methods are not unique to transport geography: they have for example been

* Corresponding author.
E-mail addresses: liang.dai@ugent.be (L. Dai), ben.derudder@ugent.be (B. Derudder), xliu6@hku.hk (X. Liu).

https://doi.org/10.1016/j.jtrangeo.2018.05.012
Received 9 November 2017; Received in revised form 7 May 2018; Accepted 7 May 2018

discussed and applied in fields as disparate as physics (Gemmetto et al., 2017), sociology (Neal, 2014), biology (Darabos et al., 2014), and computer science (Foti et al., 2011). Nonetheless, it can be noted that oftentimes the illustrative examples put forward in these domains are transport and infrastructure networks, reinforcing the broader relevance of the transport geography/network analysis-nexus. In spite of this, the adoption of the ideas developed in other scientific fields has been limited and uneven in transport geography itself (cf. Ducruet and Beauguitte, 2014).

A few comparative studies have been conducted by physicists, matching simplification methods to real-world networks (e.g. Blagus et al., 2014) and by sociologists, comparing extraction approaches for identifying statistically significant edges in social networks (e.g. Neal, 2013). However, to date there has been no systematic comparison of the relevance of different backbone extraction techniques in transport-geographical research. We therefore present a comparative analysis of key network backbone extraction techniques, discussing their practical usefulness by means of an empirical study of the Southeast Asian intercity air transport network (SAAN). This implies, of course, the potential problem of using a very specific example to derive generic claims about the usefulness of techniques. However, we believe our findings are broadly robust in the sense that this network shares common characteristics with many other transport networks and non-planar urban networks. Furthermore, we will use findings to discuss relevant applications for transport and urban geography studies more broadly, focusing on which technique works best under what circumstances and/or for what research objective.

The reminder part of this paper is organized as follows. We begin by reviewing six multidisciplinary backbone extraction techniques that are either widely adopted in geography or seem to hold potential for geographical research, and illustrate these techniques by applying them to a toy network. This is followed by a brief description in Section 3 of the data and analytical framework used to analyse the SAAN. Section 4 presents the empirical results, compares the main spatial and topological structures highlighted in each abstracted network with the original network as a benchmark, and use results to shed light on the SAAN as well as well as discussing broader applications of each method. The paper is concluded with a summary of key findings, a discussion of limitations, and possible avenues for future work.

2. Techniques for backbone extraction

2.1. Overview

Network backbone extraction has been studied in a wide range of disciplines under different names, such as network simplification (Blagus et al., 2014), sparsification (Mathioudakis et al., 2011), abstraction (Zhou et al., 2012), and reduction (Kim et al., 2011). Given that most transport networks are one-mode networks which consist of a single set of inter-connected nodes (Scott and Carrington, 2011), we will review backbone extraction techniques for one-mode networks thus excluding techniques for two-mode projections where the original networks feature connections between two different sets of nodes (Liebig and Rao, 2016).

In general, network backbone extraction techniques fall into two broad categories: "coarse-graining" and "edge removal" methods. Coarse-graining methods merge nodes sharing common attributes together and replace them by a single, new node in the abstracted network (Itzkovitz et al., 2005). The differences between approaches within this overarching logic ultimately relate to the adopted "compression" technique, i.e., the algorithm to identify communities and the rules of transformation. However, as most transport geography related research questions require retaining original nodes and edges, coarse-graining methods tend to be less appealing in our research domain.

Edge-removing techniques focus on *removing* rather than *transforming* nodes and edges; they single out the most "relevant" nodes and edges and subsequently eliminate the least significant ones. This can be achieved by edge sampling for binary networks (Blagus et al., 2014) and edge filtering/pruning for weighted networks (Bu et al., 2014). For binary networks in which 0 and 1 respectively denote the absence and presence of an edge, the abstracted network is produced by sampling the original network based on its goodness of fit to original topologies such as degree distributions, path length, assortativity and clustering coefficients (Newman, 2003). The edge sampling methods range from random node/link selection to snowball sampling, random walk, forest fire, and so forth (Lee et al., 2006). Although useful in several contexts, they are not considered in this paper since most geographical research has a vested interested in weighted networks that often combine both structural and functional aspects (Chawla et al., 2016).

In light of this, in this paper we focus specifically on the filtering/pruning techniques to extract the backbone of one-mode weighted networks. This class of methods typically employs a bottom-up strategy: they start by defining a criterion for a nodewise or edgewise examination of their "importance" or "relevance" to the network, after which the redundant edges/nodes are removed in a stepwise procedure. Different backbone extraction techniques reflect the use of different criteria for identifying node/edge "importance" or "relevance" to the network, and thus result in more or less different backbones. Understanding both the underlying logic and the empirical outcome of different techniques is therefore of the utmost importance for researchers, and in this paper we therefore explore approach and outcomes of key techniques, with a focus on transport-geographical applications. We begin our review and illustration of the different methods by using a "toy network" (Fig. 1) that is broadly in the spirit of our SAAN example. It depicts a hypothetical transport network of connections between 11 Southeast Asian cities; connections are undirected and weighted by passenger flows. Table 1, in turn, presents the backbones of the toy network as extracted by the 6 methods that will be discussed in the remainder of this section. The selection of the six methods is based on their previous adoption in transport geography research (e.g. primary linkage analysis) or because of their specific potential for geographical research in general and transport-geographical research in particular (e.g. the disparity filter algorithm).

2.2. Global weight thresholding

The most common and straightforward method is (variations to) global weight thresholding method (GWTM), a technique that only retains edges whose weights exceed a predefined threshold. The threshold can be defined as an absolute value, but also as a certain proportion of the maximum observed edge weight or the mean weight (Neal, 2013). GWTM has been extensively used since it works efficiently and produces networks that are clearly much sparser. However, most real-world networks have their edge weights unevenly distributed at multiple scales, thus making this method suffer from arbitrariness, structural bias and uniscalarity (Neal, 2014). To lessen the arbitrariness, Derudder et al. (2014) and Dai et al. (2016) propose to identify an optimal value in that the smallest network density associated with the

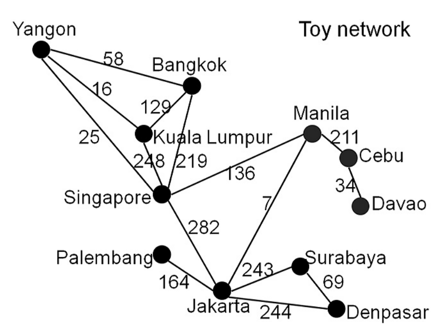

Fig. 1. Toy network (numbers next to edges represent edge weights).