Physica A 392 (2013) 4192–4201

Contents lists available at SciVerse ScienceDirect

Physica A

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

Characterizing traffic time series based on complex network theory

Jinjun Tang [a,*], Yinhai Wang [a,b], Fang Liu [c]

[a] Department of Transportation Science and Engineering, Harbin Institute of Technology, Harbin 150001, China

[b] Department of Civil and Environmental Engineering, University of Washington, Seattle, WA 98195-2700, USA

[c] Department of Energy and Transportation Engineering, Inner Mongolia Agricultural University, Hohhot 010018, China

HIGHLIGHTS

- We use complex network theory to study the traffic time series.
- Traffic time series is reconstructed using the delay embedding theorem.
- The characteristics of network are used to estimate the proper range of threshold.
- Community structure is detected according to the optimal modularity.

ARTICLE INFO

Article history:
Received 21 November 2012
Received in revised form 5 April 2013
Available online 22 May 2013

Keywords:
Complex network
Time series reconstruction
Degree distribution
Clustering coefficient
Community structure

ABSTRACT

A complex network is a powerful tool to research complex systems, traffic flow being one of the most complex systems. In this paper, we use complex network theory to study traffic time series, which provide a new insight into traffic flow analysis. Firstly, the phase space, which describes the evolution of the behavior of a nonlinear system, is reconstructed using the delay embedding theorem. Secondly, in order to convert the new time series into a complex network, the critical threshold is estimated by the characteristics of a complex network, which include degree distribution, cumulative degree distribution, and density and clustering coefficients. We find that the degree distribution of associated complex network can be fitted with a Gaussian function, and the cumulative degree distribution can be fitted with an exponential function. Density and clustering coefficients are then researched to reflect the change of connections between nodes in complex network, and the results are in accordance with the observation of the plot of an adjacent matrix. Consequently, based on complex network analysis, the proper range of the critical threshold is determined. Finally, to mine the nodes with the closest relations in a complex network, the modularity is calculated with the increase of critical threshold and the community structure is detected according to the optimal modularity. The work in our paper provides a new way to understand the dynamics of traffic time series.

1. Introduction

Traffic time series have attracted much attention due to their importance in traffic systems. Through in-depth study of traffic flow data, the characteristics and dynamics of a traffic system can be uncovered. Recently, traffic flow analysis mainly

* Corresponding author. Tel.: +86 15204638963.
E-mail address: jinjuntang@163.com (J. Tang).

0378-4371/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.physa.2013.05.012

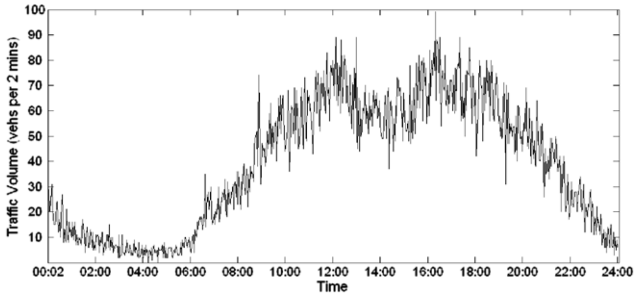

Fig. 1. Actual vehicle volume per 2 min on the highway.

includes traffic flow forecasting and nonlinear analysis. Traffic flow forecasting contains AADT forecasting [1,2], short-time forecasting (for example 2 min or 5 min) [3–5] and traffic state forecasting [6,7]. In nonlinear analysis, scholars mainly focus on adopting chaos theory to research the nonlinear characteristics in traffic flow series [8–10]. Despite abundant researching results, how to exploit the inherent connection of traffic flow series is still a significant challenge for its multiple influencing factors and dynamics.

Since the presentation of a small world network in Ref. [11], complex networks have seen a remarkable development in recent years. The publication of the study of a scale free network [12] has drawn great research interest from scholars considering real world complex networks. Complex networks contain massive nodes and edges. They represent topological structures by graphs, in which the nodes are elements and edges reflect their relations. The complex networks have been used as an effective tool to analyze social phenomena, such as the World Wide Web [13], collaboration networks [14] and transportation networks [15,16]. Furthermore, many scholars have quite recently focused on how to bridge the time series analysis and complex network. Yang et al. [17] constructed complex networks from an original stock time series, and optimized the parameters in construction by using degree distribution function. Wang et al. [18] employed a visibility algorithm to construct a complex network by using GDP series and analyzed the degree distribution, the clustering coefficient, the average path length, and community structure in a network. Gao and Jin [19] proposed a directed weighted complex network from a time series and found that time series with different dynamics exhibit distinct topological properties. Zhang et al. [20] test an extensive range of network topology statistics based on previous work and found the spatial characteristics of joint degree distribution from time series of electrocardiograms. Using statistical features of complex networks such as degree distribution and community structure provides us with a new viewpoint and a helpful method to explore the dynamic characteristics in time series. However, achievements in traffic time series combining complex networks are rarely proposed. With its different features from other time series, traffic flow data takes on a quasi-periodic nature in the long term and nonlinearity in short term. Thus, complex networks are capable of offering comprehensive statistical characteristics of the dynamics in traffic time series from a new angle.

In this paper, we use complex network to study traffic flow time series, which are measured per 2 min. Firstly, a new series is created by using a method based on phase space reconstruction, in which the C–C method [21] and G–P algorithm [22] are employed to estimate the delay time constant and embedding dimension respectively. Then, we adopt method in Ref. [17] to construct a traffic flow complex networks (TFCN) with a critical threshold value. In order to determine a reasonable value range of critical threshold, we investigate and fit the degree distribution and cumulative degree distribution by increasing the value of critical threshold. Density, clustering coefficient and adjacent matrix plot are also used to study the clustering property in the network. After generating TFCN from corresponding time series, we finally detect the community structure and find the relation between critical threshold value and modularity.

2. Methodology

2.1. Time series reconstruction

The traffic flow data are collected in Harbin city on 16th September, 2012. The measurement takes place every 2 min, in total 720 pieces of data are shown in Fig. 1.

Consider traffic time series $x_1, x_2, \ldots, x_N$, $(N = 720)$, a new sequence of phase space vectors based on the delay-coordinate embedding method [21,23] is expressed as follows:

$$\boldsymbol{Y}_i = (x_i,\ x_{i+\tau}, \ldots, x_{i+(m-1)\tau}) \quad i = 1, 2, \ldots, N-(m-1)\tau \tag{1}$$

where m is the embedding dimension, $\tau = t\tau_s$ is the delay time, t is the index lag and τ_s indicates the interval of sampling time (here, we use 2 min). Y_i means the ith reconstructed vector with embedding dimension m. Finally, we obtain a reconstructed phase space containing in total $N-(m-1)\tau$ vector points. To recover the object without distorting the topological properties of the original data, we should optimize the parameters of τ and m respectively.