Physica A 437 (2015) 235–248

Contents lists available at ScienceDirect

Physica A

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

Flow distances on open flow networks

Liangzhu Guo [a], Xiaodan Lou [a], Peiteng Shi [b], Jun Wang [b], Xiaohan Huang [b], Jiang Zhang [a,*]

[a] *School of Systems Science, Beijing Normal University, Beijing, China*

[b] *Science and Technology on Information Systems Engineering Laboratory, National University of Defense Technology, Changsha, China*

HIGHLIGHTS

- Open flow network is a directed weighted network with source and sink.
- We define flow distances on open flow networks.
- New explicit expressions of flow distances on open flow networks are given.
- We compare different distances on collected energetic food webs.
- We study the centrality and clustering of industries on the input–output network.

ARTICLE INFO

Article history:
Received 4 February 2015
Received in revised form 21 April 2015
Available online 4 June 2015

Keywords:
Open flow network
Random walk

ABSTRACT

An open flow network is a weighted directed graph with a source and a sink, depicting flux distributions on networks in the steady state mode of an open flow system. Energetic food webs, economic input–output networks, and international trade networks are open flow network models of energy flows between species, money or value flows between industrial sectors, and goods flows between countries, respectively. An open flow network is different from a closed flow network because it considers the flows from or to the environment (the source and the sink). For instance, in energetic food webs, species obtain energy not only from other species but also from the environment (sunlight), and species also dissipate energy to the environment. Flow distances between any two nodes i and j are defined as the average number of transition steps of a random walker along the network from i to j. The conventional method for the calculation of the random walk distance on closed flow networks cannot be applied to open flow networks. Therefore, we derive novel explicit expressions for flow distances of open flow networks according to their underlying Markov matrix of the network in this paper. We apply flow distances to two types of empirical open flow networks, including energetic food webs and economic input–output networks. In energetic food webs, we visualize the trophic level of each species and compare flow distances with other distance metrics on the graph. In economic input–output networks, we rank sectors according to their average flow distances and cluster sectors into different industrial groups with strong connections. Other potential applications and mathematical properties are also discussed. To summarize, flow distance is a useful and powerful tool to study open flow systems.

* Corresponding author.
E-mail address: zhangjiang@bnu.edu.cn (J. Zhang).

http://dx.doi.org/10.1016/j.physa.2015.05.070
0378-4371/

1. Introduction

A large number of studies have proven that a complex network is a powerful and useful tool to model complex systems [1–4]. However, due to the limitation of the traditional graphs for describing the complexity of various real systems, weighted networks [5,6], directed networks [7], bi-partite graphs [8], multiplexes [9,10], and temporal networks [11] have emerged as novel extensions of the conventional graphs in the past decade. Among these, an open flow network is a particular type of directed weighted network that depicts an open flow system.

Most complex systems are open, and they exchange energy and material with their environment [12]. Energy and material flows are delivered to each unit of a system by the flow network [13,14]. In open flow networks, two special nodes named the source and the sink are introduced and represent the environment. The existence of the inflows from the source and the outflows to the sink indicates the openness of the network. Because the flow system considered should be in a steady state mode, the flow network is always balanced which means that the total inflow of each node equals its total outflow, except for the sink and the source.

An energetic food web is a typical open flow network that has been studied for several years by system ecologists. The seminal work of H.T. Odum [15,16] has depicted complicated energy flow transactions between two species as an energy circuit. A group of indicators have been proposed to quantify the properties of this open flow network [17–20], and numeric common properties have been discovered [21–25]. Patten et al. proposed a systematic "Ecological Flow Analysis" method to investigate energetic flow networks [26,27].

Indeed, many basic ideas and approaches of flow analysis on energetic food webs were inherited from the economic input–output analysis method [28], which was first proposed by the famous economist Leontief [29,30]. To quantify the complex economic production processes and the interaction between different economic sectors, an input–output matrix is calculated for an economic system to represent goods flows [31,32]. Following Leontief's seminal work, Hanon introduced basic notions such as the fundamental matrix to ecology for describing the energy flows between species [33]. Therefore, an input–output matrix can also be regarded as an open flow network [34]. Money flow from the final demands compartment circulates in different sectors of an economic system and eventually flows to the value added compartment (or goods flow in an inverse direction). Thus, the value-added compartment can be regarded as the sink, and final demands can be regarded as the source. The money flow from industry i to industry j is always measured by the uniform currency unit; therefore, the total out flow from the source equals the total inflow to the sink and is identical to the gross domestic output of an economy [31,32]. Other examples of open flow networks include clickstream networks [35,36] and trade networks [37]. In summary, an open flow network is a very useful tool for depicting various open flow systems.

Distance on graph is a very useful concept [38]. The shortest path distance [39], the resistance distance [40] and the mean first-passage distance of a random walker [41–44] can reflect the intrinsic properties of the graph. However, the conventional first-passage distance on a graph is based on the basic assumption that the whole network is closed, which means the random walker cannot escape from the network; thus, the total number of walkers on the graph is conservative. Nevertheless, the open flow network is an open system. Random walkers can flow into the system from the source and flow out to the sink. The total number of walkers staying in the network can also be conservative if the flow system is in a steady state. Therefore, the traditional method for a closed system cannot simply be extended to open flow networks. It is necessary to extend the distance notions for open flow networks.

This paper is organized as follows. In Section 2, the flow distance quantities from i to j are defined. The explicit form of each flow distance is expressed in Section 2.3. Section 2.4 shows how the distance matrix is calculated on an example flow network. In Section 3.1, we apply our method to energetic food webs, visualize each species by its trophic level, and compare different distances on the food webs. The applications of flow distances on input–output networks including sector clustering and vertex centrality are introduced in Section 3.2. Finally, we give a short summary for all of the paper and the perspective of flow distances in Section 4.

2. Flow distances

In this section, we will present the definitions and calculations of flow distances. Three flow distances, namely, first-passage flow distance, total flow distance, and symmetric flow distance, are defined. They can all be expressed by using the Markov matrix of the open flow network. To obtain the final expressions, some intermediate concepts including total flow and first-passage flow are introduced.

2.1. Definitions

Consider an open flow network with N common nodes. Two special nodes are added, namely, "source" denoted by 0 and "sink" denoted by $N+1$. An $(N+2)\times(N+2)$ matrix F can be used to represent flows, and each entry f_{ij}, where $i, j \in 0, 1, 2, \ldots, N+1$, represents the flow from node i to node j. Note that the elements in the first column and the last row are all equal to zero because there is no inflow to the source and no outflow from the sink. We also define $f_{i\cdot} = \sum_{j=0}^{N+1} f_{ij}$ as the total outflow from i and $f_{\cdot j} = \sum_{i=0}^{N+1} f_{ij}$ as the total inflow to j. In our research, the flow network should be balanced, which means that $f_{\cdot i} = f_{i\cdot}$ for every node i except "source" and "sink". Particularly, we name $f_{i,N+1}$, which is the flow from i to the sink, **dissipation**.