Physica A 449 (2016) 111–125

Contents lists available at ScienceDirect

Physica A

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

Investigating the relationship between k-core and s-core network decompositions

Marius Eidsaa, Eivind Almaas *

Department of Biotechnology, NTNU - Norwegian University of Science and Technology, N-7491 Trondheim, Norway

HIGHLIGHTS

- A systematic study of the transition between k-core and s-core analysis is proposed.
- The node content in innermost k-cores and s-cores is similar for many network topologies and link-weight combinations.
- For scale-free networks with positively correlated link weights, the innermost s-cores are robust to link-weight discretization.

ARTICLE INFO

Article history:
Received 13 August 2015
Received in revised form 2 December 2015
Available online 6 January 2016

Keywords:
Complex networks
Weighted networks
Network decomposition
k-core analysis
k-shell analysis
s-core analysis

ABSTRACT

Network decomposition methods, such as the much used k-core analysis, are able to identify globally central regions of networks. The decomposition approaches are hierarchical and identify nested sets of nodes with increasing centrality properties. While most studies have been concerned with unweighted networks, i.e. k-core analysis, recent works have introduced network decomposition methods that apply to weighted networks. Here, we investigate the relationship between k-core decomposition for unweighted networks and s-core decomposition for weighted networks by systematically employing a link-weight scheme that gradually discretizes the link weights. We applied this approach to the Erdős–Rényi model and the scale-free configuration model for five different weight distributions, and two empirical networks, the US air traffic network and a Facebook network. We find that (1) both uniformly random and positively correlated link-weight distributions give rise to highly stable s-core decompositions with respect to discretization levels. (2) For negatively correlated link-weight distributions, the resulting s-core decomposition has no similarity to the k-cores. Since several combinations of network topology and link-weight distributions give rise to a core-structure that is highly similar to the full s-core for a large range of link-discretization levels, it is possible to significantly speed up the numerical s-core analysis for these situations.

1. Introduction

A large array of methods exists for the analysis of complex networks where all of the links have the same (unit) strength [1–4]. The aim of applying these methods is to identify nodes, or sets of nodes, with special properties. In particular, there is great interest in identifying nodes that are central in the structure of a network, or important for a spreading process that takes place on the network. Notable among these methods are the core-decomposition analyses, where k-core analysis [5] is the most basic representation. Typically, core-decomposition methods seek to find nodes that are

* Corresponding author.
E-mail address: eivind.almaas@ntnu.no (E. Almaas).

http://dx.doi.org/10.1016/j.physa.2015.12.115
0378-4371/

centrally important to the network by iteratively peeling off less-central nodes, leaving a sequence of increasingly more interconnected sets (cores) of nodes.

The k-core network decomposition has been useful for the analysis of a wide range of systems, spanning biological networks [6,7] to technological systems, such as the Internet [8], and transportation networks [9]. The k-core analysis has also been employed in addressing the demanding task of visualizing large networks [10]. Motivated by the challenge of identifying nodes that are influential spreaders in a network, several alternative versions of core analysis and network decomposition have recently been proposed [11–16].

In many network representations of systems, it is natural to associate a strength, or weight, with each link. When taking the possibility of unequal link-strength into consideration, important aspects of the system are highlighted that address the confluence of network topology and the (often dynamic) processes giving rise to link weights [17–19]. While k-core methodology solely takes a network's topology into consideration, i.e. which pairs of nodes are connected, recent works have suggested generalizations of the core decomposition approach to weighted networks [20–22]. These methods are, in short: a k-shell generalization, $W_{k-\text{shell}}$, where the weighted degree (or the $(\alpha+\beta)$th root of the product of k_i^{α} and $(\sum_j w_{ij})^{\beta}$) is defined as the node-pruning measure [20], a k-core generalization using node strength s instead of k as node removal threshold [21], and a weighted k-shell method where the link weight is defined as $w_{ij} = k_i + k_j$, and the weighted node degree is set to be $k_i^{w} = \alpha k_i + (1-\alpha)\sum_j w_{ij}$, with α as a tuning parameter [22]. Thus, these weighted network decomposition methods incorporate the consequences of unequal link weights in the network analysis of node-centrality.

In this article, we study properties of the s-core network decomposition method [21], and in particular, how it relates to and differs from the traditional k-core analysis [5] through a link-weight discretization scheme. Since link-weights often are non-degenerate, leading to node strength values that are non-degenerate, the s-core analysis may use considerably more computer time than the k-core analysis which leverages large levels of node-strength (i.e. degree) degeneration. However, if discretized versions of the link weights give rise to s-core structures that are highly similar to those resulting from the non-discretized version, s-core analysis will be computationally about as fast as k-core analysis. Note that, when the s-core approach is applied to a network where the link weights all have the same value, it will produce a core structure that is identical to that of the k-core method [21].

We will focus our investigation on the relationship between the k-core and the s-core network decomposition methods by employing a link weight scheme that gradually discretizes link weights, and thus makes it possible to gradually transition from k-core to s-core. For each of the investigated network topologies, link weight distributions, and chosen weight-discretization levels, we calculate the s-core decomposition of the system and contrast with results corresponding to that of the k-core decompositions for an unweighted version of the same network. We have chosen two different random network models: Erdős–Rényi model [23–25] and the configuration network model [26–28] with scale-free connectivity distribution, since these two models capture the essence of the majority of empirical networks studied thus far. For each of the models, five link-weight distribution schemes were chosen: Two of the distributions have link weights that are correlated with network topology, and three are uncorrelated random distributions. Additionally, we study two empirical, weighted networks: US air traffic network for 2010 to 2014 [29] and a Facebook reply network [30–32].

2. Method

In the following, we will use the notation from Ref. [21]: A network consisting of N nodes and M links is described by the adjacency matrix $A = [a_{ij}]$, where $a_{ij} = 1$ if nodes i and j are neighbors, and 0 otherwise. The degree of a node is the number of its nearest neighbors, i.e. $k_i = \sum_j a_{ij}$. Weights can be assigned to the network by introducing a number w_{ij} defining the strength of the link between nodes i and j. The node strength is defined as $s_i = \sum_j a_{ij}\left|w_{ij}\right|$. The k-core is defined as the remaining subset of nodes i with degree $k_i > (k-1)$, when all nodes i with degree $k_i \le (k-1)$ have been removed recursively. The s-core is the weighted generalization of the k-core, where s is a node strength threshold value $s \in \mathbb{R}^+$.

Similar to k-core definition, the s-core of a weighted network consists of the subset of nodes i with strength $s_i > s$. This is achieved by iteratively removing all nodes i with strength $s_i \le s$. In analogy to the indexing of k-cores, we rank the s-cores as follows: The s_{n+1}-core has the threshold value $s_n = \min_i s_i$, where node $i \in s_n$-core network. The s_0-core is defined to contain the entire network, including nodes with no neighbors. The s_1-core is defined to contain all nodes with nonzero strengths, analogous to the $(k=1)$-core, i.e., $s_0 \equiv 0$ and $s_1 = \min_i s_i, s_i \neq 0$. Instead of using the actual strength threshold value s_n to describe a distinct s-core, it is customary to use the integer-valued s-core index n for simplicity. Note that if $w_{ij} = w$, a constant, s-core analysis will yield a core decomposition that is identical to that of k-core analysis.

To describe the node peeling nature of a core decomposition, it is convenient to define the k- and s-shell. The k-shell is defined as the subset of nodes that are members of the k-core, but not the $(k+1)$-core. Similarly, the s_n-shell is defined as the subset of nodes that are contained in the s_n-core, but not in the s_{n+1}-core. In case of continuous weight distributions, where few nodes (if any) share the same strength values, we expect the majority of s-shells to contain a small number of nodes. Consequently, we expect to find $n_{\max} \sim \mathcal{O}(N)$. For discrete weight distributions, many nodes may take the same strength value, resulting in larger s-shells and a smaller $n_{\max}/N$ ratio.

We will assume that all weights are finite, making the weight distributions scalable so that $w_{ij} \in (0, 1]$. Note that the s-core decomposition is invariant under this monotone scaling. Furthermore, we discretize the link weights by dividing the interval $(0, 1]$ into m bins. As an example, consider the scenario where we make approximations of the link weights w_{ij} by