Coreness Tunable Network Model

Yifan Wang*
Institute of Transportation Systems Science and Engineering, Beijing Jiaotong University, Beijing 100044, China
*825851343@qq.com

Abstract. Kitsak et al argued that nodes dwell in diverse network shells by a k-core decomposition process show more reliable identification for nodal importance which had attracted more and more attentions in different domains. But one seldom focuses on the distribution of node numbers (DNN) in different shells of a network, experiment results show regular characteristics. While, the existing theoretical network models, such as BA scale-free model, WS Small-world model, ER random network model et al cannot reproduce the features. To fill this gap, a group of coreness tunable network (CTN) models are proposed, in which the coreness of each node is totally controllable. The CTN has a similar network performance compared to real-world by counting basic static geometric features and spreading performance under SIR model. Our CTN models are providing a theoretical framework to deepen humans' understanding of the coreness structure and function of complex networks.

Keywords: Coreness Tunable Network Model; Shell Distribution; SIR Model.

1. Introduction

A identification of core in a network by core decomposition process was firstly introduced by Seidman[1] to study tightly-knit groups in social networks. The *k*-core may be obtained by the following way. Iteratively remove the all nodes with degree k until there is no node left with degree k in a network. The removed nodes, along with corresponding links form a k shell with index equaled k. This decomposition helps one to describe the complex topologies of real-world networks and was applied to a number of real-world networks for example the Internet, the WWW, protein networks, etc. which was turn out to be an important tool for visualization of complex networks and interpretation of cooperative processes in them[2].

Recently, Kitsak et al[3,4] find that the nodes in different shells (coreness) of a network can identify the influential nodes more accuracy compared by the intuitive statistics indicator degree. Owing to its accuracy for identifying influential nodes and low computational complexity, coreness can be found many applications in many real networks[5-8] and improved for more reliable and precise ranking methods[4,7,9-16].

Methods for identifying spreaders using cores were extended to dynamic networks in Miorandi and Pellegrini[15] and core decomposition in general was extended to weighted netwotks in Eidesaa and Almaas[17].

Although core decomposition has become an important and widely used tool as a descriptive summary statistic of the network, it is a statistic for which there does not exist a theoretical network model[17]. Last year, Karwa had researched the shell distribution in random networks. However, coreness cannot be used in some classical modeled networks, such as BA networks [18] and tree-like networks, where the coreness values of all nodes are the very small and indistinguishable.

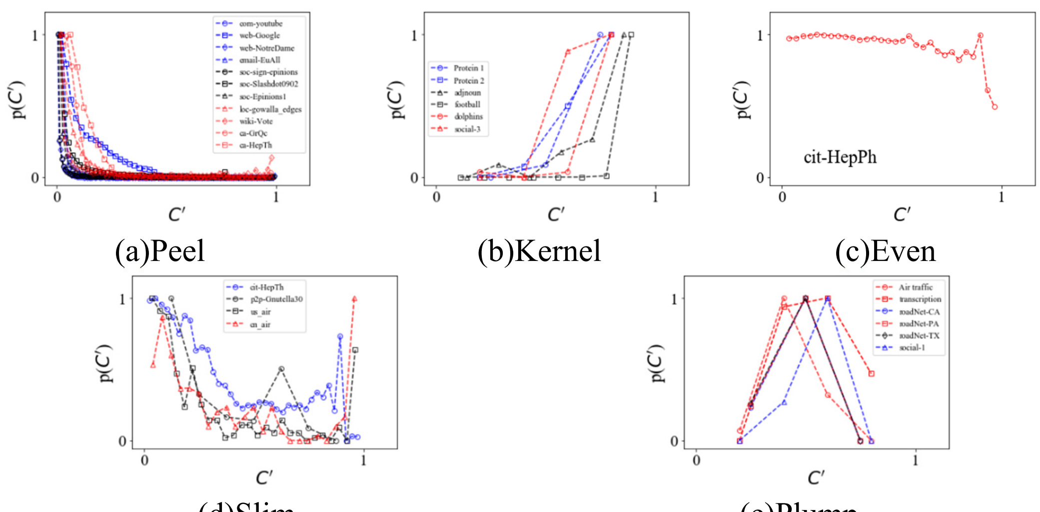
In this paper, by investigating lots of real-networks we propose a shell connection mechanism (SCM) which can control the coreness of per node in the network. Experimental results show that the SCM networks can generate all topological types of model networks by reproducing the arbitrary shell distribution. And the generated network has a similar dynamic performance compared to the real network.

2. Shell Distribution

We consider that the network topological structure to the main core layer (30% shells), second core layer (30% shells) and marginal layer (30% shells) to form a multi-level "core-periphery" topological mode. To study the core decomposition process, we apply the k-core (also called k-shell) decomposition analysis. This process assigns an integer index or coreness, k_s, to each node, representing its location according to successive layers (k shells) in the network and the network can be viewed as the union of all k shells. Counting the number of vertices in each shell, which is namely shell distribution, as a sufficient statistic. Empirical studies on real-world networks and the previous theoretical networks show two types of phenomena. To compare the shell distribution of different networks, we have

$$p(c) = \frac{N(c) - N(c_{\min})}{N(c_{\max}) - N(c_{\min})} \tag{1}$$

where $N(c)$ is the number of the nodes with coreness c. Fig.1 shows that shell distribution of real networks can be classified into five types. (a) "Peel" networks. Nearly half of real-world networks' shell distributions show that most of the nodes are located at the periphery of the network. (b) "Kernel" networks. Contrary to "Peel" networks, the nodes of the 6 networks are mostly distributed at the inner core of the network. (c) "Even" networks. In addition to the largest shell layer of the network, the nodes are distributed evenly among the shell layers in the network. This distribution occurs in cit-HepPh network, a paper citation network, which can be considered as a "Even" network. (d) "Slim" networks. The nodes are mainly concentrated in the periphery and inner of the network, and only a small number of nodes in the middle shells. (e) "Plumpy" networks. Opposite to "Slim" networks, most nodes are distributed in the middle shells. This type of networks is generally road network, aviation network and so on.

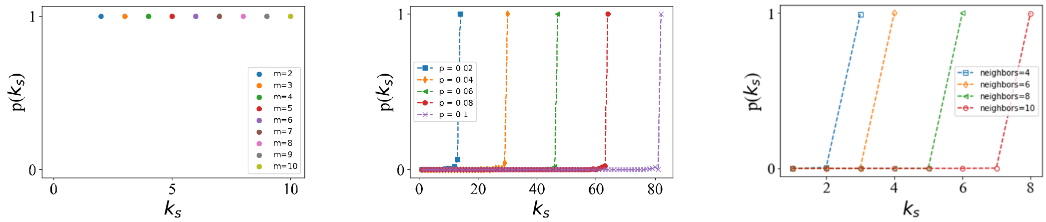


(a)Peel (b)Kernel (c)Even

(d)Slim (e)Plump

Figure 1. Shell distribution of 28 real networks.

However, we select BA scale-free network, WS small world network and ER random network as the previous theoretical networks to k-shell decompose. The results of the previous theoretical networks are shown in Fig.2. In (a) BA scale-free network [18], multiple experiments were performed under different parameter m which is the number of the edges of the new node. The results show that all the nodes are distributed at the largest shell layer which equal to m. In (b) WS small world network [19], experiments are conducted under different k which is the number of neighbors connected by each node. Almost all nodes are located in the maximum shell. In (c) ER network [20], nodes are always clustered at the inner core of the network under different probabilities for edge creation. From what has been say discussed above, these previous theoretical networks do not work in the study of coreness because they cannot reproduce the shell distributions of the real-world networks. So, a new applicable model network is needed.



(a)BA scale-free network (b)ER random network (c)WS small world network

Figure 2. Shell distribution of three theoretical networks.

3. Coreness Tunable Network MODEL

Usually, the k-core decomposition of network is used to understand the structure of the network, or identify the most efficient spreaders. In this study, we focus on building a theoretical model for studying the properties of coreness. From the above, we know that the k-core decomposition is a process of decompose the network according to shell index and gives nodes with coreness. But the inverse process of it is a process of generating network. If given a shell index sequence of per node, we can capture the network's core structure and build a new network with same shell sequence by the inverse process of k-core decomposition. So, we proposed a group of coreness tunable network models based on it.

When building CTN models, there are three aspects should be considered. The first is the way nodes joined the network. The model network can add one node at a time, or multiple nodes at a time. To simplify the experiment, we consider adding one node at a time or adding a set of nodes with the same coreness at a time. The second is the mode new nodes connect to the network. The new node that can connect to the previous shell layer, or connect to all the previous shell layers, or connect all nodes in the network. The last one is the number of nodes in each shell. It needs to be clear that the quantitative limitation of the nodes in each shell whether affect the coreness of per node in the network. Based on the above problems, a set of CTN models are constructed to verify whether these problems affect the coreness of nodes in the process of network generation.

As Fig.3 shown, given a desired coreness series $(c_1 = c_{\max} \geq c_2 \geq \cdots \geq c_{N-1} \geq c_N)$, CTN model constructs a complete network of $(c_{\max} + 1)$ nodes at the beginning. All nodes in the complete network have the same coreness $c_{\max}$. Then adding the rest nodes follow the descending order of the coreness series. At each time step, a new node *i* is randomly connected to c_i existed nodes in the current network.

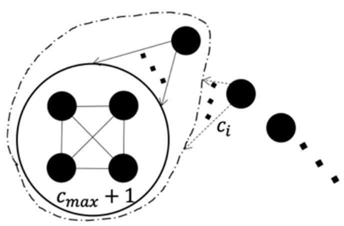
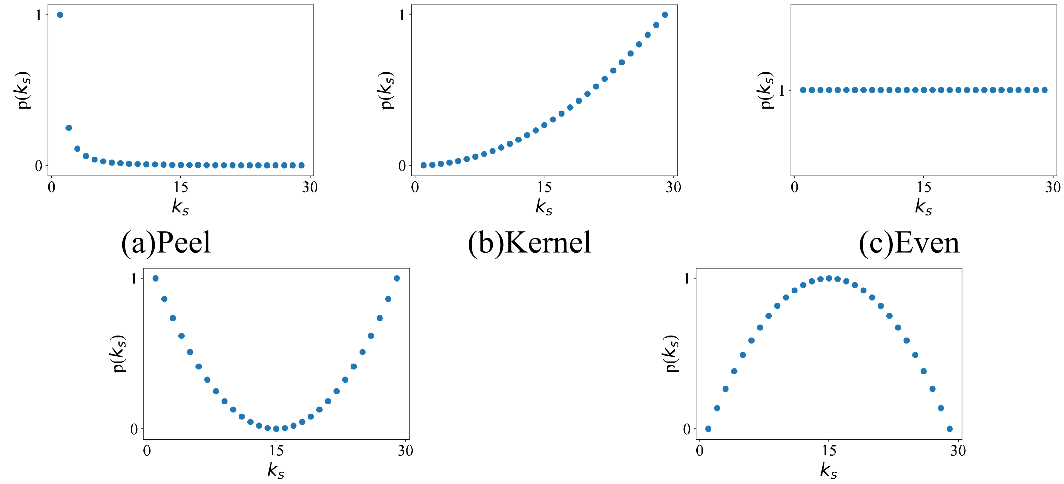


Figure 3. Coreness tunable network model.

4. Reproduce Shell Distribution of Real Networks

Fig.4 shows the perfect five types of shell distribution produced by CTN model.



(a)Peel (b)Kernel (c)Even

(d)Slim (e)Plumpy

Figure 4. Shell distributions for the theoretical networks by the CTN model

5. Statistical Characteristics of CTN Model

To comprehensively understand CTN model networks and their corresponding real networks with the same coreness series, we choose four typical real networks, and compare the structural properties in Table 1. Structural properties include number of nodes (*N*), number of edges (*E*), average shortest path length (*L*), clustering coefficient (*C*), average degree ($\bar{k}$), maximum coreness index ($c_{\max}$). Dynamic properties include the epidemic threshold (β_c)and average size of the infected population (<<M>>). CTN model firmly constrains that the number of nodes and the coreness series should be accurately the same with its corresponding real network, but the number of edges is required approximately. Each of the four CTN model networks has 100 realizations, and the properties in Table 1 are averaged over them. We can see that the *E* and $\bar{k}$ of the CTN model networks are higher than those of the real-world networks. However, the *L* and *C* of the CTN model networks are lower than those of the real-world networks, which means that although the density of the CTN model is large, there is no obvious clustering effect.

Table 1. Properties of four typical real networks and their corresponding CTN model networks.

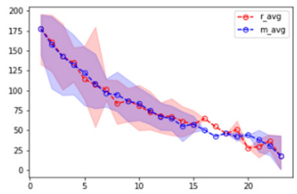
Network	*N*	*E*	$\bar{k}$	*L*	*C*	$c_{\max}$	β_c	<<M>>
China air	196	1580	16.12	2.19	0.68	23	0.023	1.21×10^{-2}
CTN-China air	196	1591	16.23	2.45	0.43	23	0.028	1.05×10^{-2}
us_air	332	2126	12.81	2.73	0.63	26	0.021	7.1×10^{-3}
CTN-us_air	332	2162	13.02	2.63	0.42	26	0.026	6.41×10^{-3}
web-google	1299	2773	4.27	6.47	0.35	17	0.088	1.72×10^{-3}
CTN- web-google	1299	3974	6.12	2.56	0.34	17	0.010	2.89×10^{-2}
ca-GrQc	5242	14484	5.53	6.05	0.53	43	0.059	6.67×10^{-4}
CTN-ca-GrQc	5242	19764	7.54	3.59	0.3	43	0.003	2.0×10^{-2}

6. Dynamic Characteristics of CTN Model

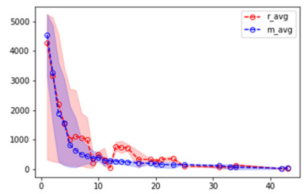
To see whether the dynamic characteristics of CTN model generated network can reflect the real, we study the standard the susceptible-infectious-recovered (SIR) model in which the reflection is confirmed using comparing the influence of node i in both networks after the dynamics over 200 independent runs, each of which begins with node i as the sole infected seed. We denote the probability that an infectious node will infect a susceptible neighbor as β. In our simulations, we chose a little β value equaled β_c , where $\beta_c = \langle k \rangle / \left(\langle k^2 \rangle - \langle k \rangle \right)$ is the epidemic threshold of the network determined by the heterogeneous mean-field method[21].

We make a further analysis of the similarity of dynamic characteristics between real networks and CTN model networks. We comparing the average and extreme values of the node influence between CTN model networks and their corresponding real networks with the same coreness series. As shown in Fig.5, in most cases, the ranking of nodes in each shell of CTN model is more stable than that of the real and the mean values of the rankings are very approximate.

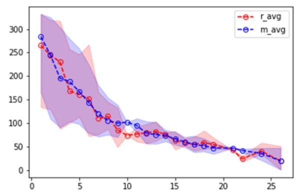
A CTN model network is constructed according to the coreness series of the real, and the size of the coreness, an inaccuracy statistic, reflects the distance from nodes to the core of the network. So, it is incorrect to compare the correlation of two kinds of networks in the dynamic characteristics using the comparison of the single node infection ability. So, we consider the sequence of infection capacity of nodes and results show that CTN model has dynamic characteristics approximate to real-world networks.



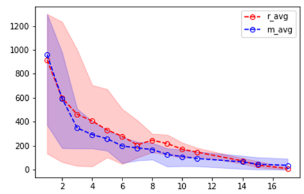
(a)China air



(b)USA air



(c)Web-google



(d)Ca-GrQc

Figure 5. Comparison of each shell capability between CTN model and real network

7. Summary

Coreness have been widely used to study and summarize networks. In this paper we study the k-core decomposition of a network with an eye towards reproducing the shell distribution. We find that the shell distribution of the real-world networks can be divided into five types according to their shapes. In most cases, the real networks belong to “peel” networks in which most of nodes are located on the fringe of the networks. However, the classical theoretical models such as BA scale-free network, ER network and WS small world network, cannot reproduce these distributions. So, in order to study coreness better, we need to construct a new model to fill in the gaps and then do some further research on the character of the model. Based on the discussion about whether the way nodes join the network, the mode new nodes connect to the network, and the number of nodes in each shell affect the coreness of nodes when the model is generated, we structure six models to test their impact on the coreness of nodes. We find that these three factors do not affect the coreness and five model networks with different shell distributions are generated by the model 6 which is closest to the real network in all models. Then we discuss the relationship between the statistical characteristics and dynamic characteristics of the model network and that of the real network in the comparison of networks of different sizes. In terms of statistical characteristics, our model can well reflect the nature of the real network in the perspective of coreness. In terms of dynamic characteristics, the model has the properties similar to the real network. The proposed model bridges the gaps between the real-world network and theoretical network in the respect of coreness and may have potential applications in researching the relationship among different network topologies under the same coreness sequence. And providing a theoretical framework to deepen humans' understanding of the coreness structure and function of complex networks.

References

[1]. S.B. Seidman, Network structure and minimum degree ☆, Social Networks, 5 (1983) 269-287.

[2]. S.N. Dorogovtsev, A.V. Goltsev, J.F.F. Mendes, k-core organization of complex networks, Phys. Rev. Lett., 96 (2006) 4.

[3]. M. Kitsak, L.K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H.E. Stanley, H.A. Makse, Identification of influential spreaders in complex networks, Nat. Phys., 6 (2010) 888-893.

[4]. J. Bae, S. Kim, Identifying and ranking influential spreaders in complex networks by neighborhood coreness, Physica a-Statistical Mechanics and Its Applications, 395 (2014) 549-559.

[5]. S. Wuchty, E. Almaas, Peeling the yeast protein network, Proteomics, 5 (2005) 444–449.

[6]. S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt, E. Shir, A model of Internet topology using k-shell decomposition, Proc. Natl. Acad. Sci. U. S. A., 104 (2007) 11150-11154.

[7]. A. Zeng, C.J. Zhang, Ranking spreaders by decomposing complex networks, Phys. Lett. A, 377 (2013) 1031-1035.

[8]. S. Pei, L. Muchnik, J.S. Andrade, Z.M. Zheng, H.A. Makse, Searching for superspreaders of information in real-world social media, Scientific Reports, 4 (2014) 12.

[9]. L. Lu, T. Zhou, Q.M. Zhang, H.E. Stanley, The H-index of a network node and its relation to degree and coreness, Nat. Commun., 7 (2016) 7.

[10]. A. Namtirtha, A. Dutta, B. Dutta, identifying influential spreaders in complex networks based on kshell hybrid method, Physica a-Statistical Mechanics and Its Applications, 499 (2018) 310-324.

[11]. Y. Liu, M. Tang, T. Zhou, Y. Do, Identify influential spreaders in complex networks, the role of neighborhood, Physica a-Statistical Mechanics and Its Applications, 452 (2016) 289-298.

[12]. Z.X. Wang, C.J. Du, J.P. Fan, Y. Xing, ranking influential nodes in social networks based on node position and neighborhood, Neurocomputing, 260 (2017) 466-477.

[13]. Z.X. Wang, Y. Zhao, J.K. Xi, C.J. Du, Fast ranking influential nodes in complex networks using a k-shell iteration factor, Physica a-Statistical Mechanics and Its Applications, 461 (2016) 171-181.

[14]. L.Y. Lu, D.B. Chen, X.L. Ren, Q.M. Zhang, Y.C. Zhang, T. Zhou, Vital nodes identification in complex networks, Phys. Rep.-Rev. Sec. Phys. Lett., 650 (2016) 1-63.

[15]. D. Miorandi, F.D. Pellegrini, K-shell decomposition for dynamic complex networks, Proceedings of the International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, 2010, pp. 488 - 496.

[16]. M. Eidsaa, E. Almaas, s-core network decomposition: a generalization of k-core analysis to weighted networks, Phys Rev E Stat Nonlin Soft Matter Phys, 88 (2013) 062819.

[17]. M. Eidsaa, E. Almaas, s-core network decomposition: A generalization of k-core analysis to weighted networks, Phys. Rev. E, 88 (2013) 9.

[18]. A.L. Barabasi, R. Albert, Emergence of Scaling in Random Networks, Science, 286 (1999) 509.

[19]. D.J. Watts, S.H. Strogatz, Collective dynamics of |[lsquo]|small-world|[rsquo]| networks, Nature, 393 (1998) 440-442.

[20]. P. Erdős, A. Rényi, On Random Graphs I, Publicationes Mathematicae, 4 (1959) 3286-3291.

[21]. C. Castellano, R. Pastor-Satorras, Thresholds for Epidemic Spreading in Networks, Phys. Rev. Lett., 105 (2010) 4.