Node clustering in complex networks based on structural similarity

Feng, D.; Li, M.; Zhang, Q.

Physica A: Statistical Mechanics and its Applications

Volume 658, 15 January 2025, 130274

ex networks based on structural

ɔdes based on the k-means

ructural similarity of the

l to community detection.

de clustering is a method that identifying nodes with the same function or mmunity detection. In this work, based on the structural similarity of nodes de clustering is proposed. This method can easily divide the hub nodes and heral structure into two sets. We also find that the changes in the number of rules that guide the growth of the network under different conditions. e Erdős–Rényi model, the cluster of nodes is homogeneous. When the nodel, the peripheral nodes are the majority, and this trend will not change v that the clustering of nodes in the networks based on the nodes' structural h on the structural analysis of complex networks.

Next

plex networks

l-world' [1]and 'scale-free' [2] have been found in networks, triggering x networks are not only a new kind of models for physics or mathematics [3],

[4], it also a new tool that can be applied in different research fields, such as biology [5], [6], neuroscience [7], [8], financial systems [9], [10], and even social sciences [11], [12]. Over time, various researchers have invested significant efforts in understanding the deep implications of complex networks, covering explorations of node centrality [13], [14], [15], self-similarity [16], [17], [18], statistical mechanics properties [19], [20], [21], [22], [23], and dynamic behavior [24], [25], [26], [27]. These studies delve into the rich connotations of complex network structures, providing important theoretical foundations for

onents of complex network research. In the analysis of complex network in the topological structure is a core problem. These structural information], [30], quantifying structural similarities between nodes [31], [32], and k [14], [33], [34], [35]. Among those researches, the detection of community role.

ysis of networks, network community detection is the automated discovery of re similar features or roles [36]. This involves partitioning nodes in the groups are densely connected while nodes between groups are sparsely e in complex networks builds a bridge to understanding the structural ucture to the macro characteristics. Thus, studying the structural properties of nodes can reveal the organizational principles and operational functions of n methods have been proposed [37], including clustering [38], divisive 40], [41], [42], dynamic algorithms [43], [44], and statistical inference [45], ods try to find the groups of nodes that are strongly connected. Only a few s that share similar features or roles, which is important for the macroscopic

ity structure in complex networks is analogy with the clustering algorithm in ormation of each data point (nodes) and their neighbors and try to find the detecting of the community in complex networks can be treated as finding the nd gather those nodes with the same structural information into a same set of the order of those evidential structural information carried by those nodes

lustering algorithm and nodes' structural information to cluster the nodes in ormation about the whole network. Traditional computer clustering methods relocation [52], [53], and density-based clustering [38], [54]. When combining in complex networks, the k-means++ method is easy to apply. Compared to the dispersed when selecting initial cluster centers, which helps avoid the tter robustness and stability. Thus, we chose K-means++ to cluster the nodes milarity, which is defined by the topological similarity of the local network lative entropy [31], [55], [56].

ion 2 introduces the preliminary work on calculating the structural relative roposes a node clustering method based on network structure and relative es to guide the growth of core–periphery and randomly distributed seed e networks using the proposed method and provide results and analysis.

des' similarity is an important metric, which is crucial for many research h as node classification, network community structure detection, and network nodes' similarity quantification based on the topological structure similarity ode is a unique one [31]. This method is based on the local network of each dges between them, and each node itself is also included in its local network,

probability sets, calculating relative entropy, and quantifying node similarity. re N represents the set of nodes in the local network and D represents the set ee in the entire network is defined as D_{max}. $P(i)$ denotes the probability set

for node i. Considering that each node's local network includes its neighbors and itself, The size of the probability set, denoted as m, is as follows:

$$m = D_{max} + 1. \tag{1}$$

The probability set associated with node i is defined as:

] . (2)

ent in the probability set P(i) is determined based on the degree set D in the owever, the degrees of most nodes are smaller than D_{max}. Consequently, when nents in the probability set are set to zero. The definition of p(i, k) under this ee of node i, D(k) is the degree of node k in the local network$L_i\ (N, D)$.

(3)

ects the relative entropy and the accuracy of the similarity measure. Before the probability set, we first order the elements, denoting the sorted set as llows:

(4)

he local structure of the node 3. Network A is shown in sub-figure (a), while 3 in network A.

may be equal to 0, and when the value of P '(j, k) is equal to 0, there will be we redefine relative entropy as

(5)

(6)

presents the difference in their local structure distributions. The relative gnificant difference in the local structure between the corresponding nodes. es a higher similarity in the local structure between nodes. Therefore, based ir of nodes, we will have the incidence matrix R of the network as

(7)

of nodes, but the relative entropy is not symmetric. Therefore, the associated

(8)

pair of nodes, we can establish an association matrix R to describe the

similarity relationship between nodes in complex networks. The entries in this association matrix reflect the relative entropy values between nodes, depending on their local structures. If the relative entropy value between nodes is large, the corresponding value in the association matrix will also be large, and vice versa. Therefore, by observing the values in the association matrix, we can understand the degree of similarity between nodes in the network.

milarity

thod for measuring node similarity in complex networks based on relative s section, we propose a new node clustering method based on the similarity r method draws inspiration from and improves upon the k-means++ clustering gorithm relies on the coordinate representation of data points, whereas our similarity, eliminating the need to directly access the coordinates of data ole. To address this issue, we enhance the k-means++ algorithm and apply it to ween nodes in complex networks using local network information and the nal k-means++ algorithm, which employs Euclidean distance to evaluate the ids, our method better captures the intrinsic structure and patterns of the r updating rules, distance calculation between sample points and cluster ncing the algorithm's performance and stability. The algorithm process is as

reviously proposed methods of relative entropy and node local structure, we se it as a measure of similarity denoted as $D(i,j)$, where r_{ij} represents the

(9)

ion process introduced earlier, we obtained the correlation matrix R of each element represents the similarity between corresponding nodes. These $D(i,j)$ in the subsequent clustering algorithm.

tions, we will use the first five nodes of Network A as an example to detail the e values of the distance matrix are shown in Table 1.

domly select a node as the initial cluster center c_1. In the subsequent process quared distance from each node x in the dataset to the currently selected been selected, we will compare the distance of each node to all the selected ng the squared distance. Subsequently, we construct the probability t cluster center based on the calculated squared distance values. The specific

(10)

ance from node x' to its nearest centroid, while X is the set of all nodes in the es that are farther from the existing centroids have a higher probability of ased on the constructed probability distribution, we randomly select the next cess until k cluster centers are selected: $C = \{C_1, C_2, \ldots, C_k\}$. The detailed lgorithm 1.

ork A and set $k = 2$. First, we randomly select a node as the initial cluster ate the squared distances from each of the other nodes to Node 1:

$$.399, \quad D_{14}^2 = (0.782)^2 = 0.611,$$

$$.774$$

According to the K-Means++ algorithm, the probability of a node being selected as the next center is proportional to the square of its distance. The probabilities of each node being selected as a center are calculated as follows:

$$P_2 = \frac{0.251}{2.774} \approx 0.0905, \quad P_3 = \frac{0.399}{2.774} \approx 0.1439, \quad P_4 = \frac{0.611}{2.774} \approx 0.2203,$$
$$P_5 = \frac{1.513}{2.774} \approx 0.5453.$$

Reading Assistant

Questions you could ask:

Actions you could take:

Sign in to unlock the full response and ask your own questions.

Sign in

re more likely to be selected as cluster centers. According to the calculation, as the next cluster center, updating the set of cluster centers to $C = \{1, 5\}$.

ances from the remaining nodes (Nodes 2, 3, and 4) to each of the current nearest center. Based on these distances, we then construct the d distances from Node 2 to Nodes 1 and 5 are as follows:

.031

as the nearest center for Node 2. Similarly, we calculate the squared distances mine their respective nearest centers. Based on the squared distances from bability distribution that gives nodes further from any center a higher ter. Using this probability distribution, we then choose the next cluster center,

d cluster centers

eger(1, length(D))

$1 : i-1], :]$

ension=1)

(min_dists^2)

nple(1 to length(D), 1, true, probabilities)

ce from each data point to each cluster center using the distance matrix D, and xt, for each cluster C_j, identify the data point that best represents the cluster ation of the iterative update algorithm is shown in Algorithm 2.

rinciple of minimizing the internal distances of the clusters. Specifically, we minimizes the average distance to all other points in the cluster. This can be blem:

er C_j, compute its distance to the other nodes using the

eighbor of node x_i, i.e., the node with the minimum distance to sing the following formula:

(11)

um value", which means finding the node k that minimizes the node that is nearest to node x_i within cluster C_j, which is used

3. **Count Frequency**: Record the nearest neighbor indices k_j of all nodes x_j. Calculate the occurrence frequency of each index in the array $\{k_1, k_2, \ldots\}$ and select the node with the highest frequency as the new cluster center. If multiple nodes have the same frequency, randomly select one as the new cluster center to ensure that the newly selected center best represents the current category.

; until the cluster centers no longer change significantly or until this point, the clustering process is complete, and all sample

ization(D, K)

ırest centers based on dists

ters) **do**

do

ned to cluster i

clusterNodes] + I

inimum distances from subDists

es of minDistIndex

most occurrences in counts

')

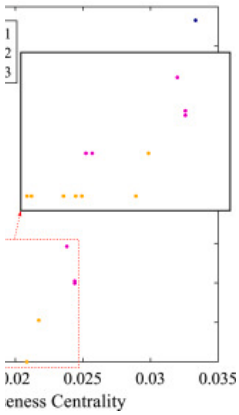
iteration ≥ max_iterations **then**

cluster centers set as Node 1 and Node 5. Based on the distance matrix D, we various cluster centers and assign the data points to the nearest cluster center. n node 2 to node 1 is 0.501, while the distance to node 5 is 0.176, thus node 2 is r 5. Continuing to calculate the distances of nodes 3 and 4 to the initial cluster = $\{1\}, C_2 = \{2, 3, 4, 5\}$.

For the update of the cluster centers, we find the nearest neighbor nodes for each node in $C_2 = \{2, 3, 4, 5\}$ based on the distance matrix D. The distance between node 2 and node 3 is 0.051, making node 3 the nearest neighbor node of node 2. Similarly, the nearest neighbor node of node 3 is node 2, the nearest neighbor node of node 4 is node 3, and the nearest neighbor nodes of node 5 are node 3 and node 4. Next, we count the frequency of each node appearing as the nearest neighbor node and select the node with the highest frequency as the new cluster center. The statistical results show that node 3 has the highest frequency

ıd 5), thus node 3 is chosen as the new cluster center. The updated cluster tered at node 1, while C_2 is updated to be centered at node 3. This process will e or the maximum number of iterations is reached. The final clustering result is

ully clustered it into three classes. To illustrate the clustering results more ıre graph and scatter plot in Fig. 2, where the horizontal and vertical axes ntrality, respectively. By observation, we can clearly see the distribution of the nodes in both the network's structure and functionality.



ɔsition of nodes in network A with different clusters is shown in the left sub- ılts is shown in the right sub-figure. In the right sub-figure, the horizontal axis is represents betweenness centrality.

fewer connections in the network and play relatively marginalized roles, with e the core nodes, which hold the most critical positions and functions in the hin the network and exhibit significant closeness centrality and betweenness the network and are crucial for its stability and functionality. We also identify ipheral and core nodes in the network. These nodes act as bridges between l exchange of information within the network. From the test of our method on reveals the distribution and roles of different types of nodes in the network, r research and analysis of network structure.

section shows that the clustering of the nodes based on the k-means++ ıster the nodes in different groups, and each group of nodes has its own vork's topological structure will be illustrated by the change of the clusters ıteresting question needs to be answered: if the change of the clusters of nodes ıat rules guide the network's growth? In other words, can the change in the be used as a general tool for the structural analysis of complex networks?

ıct two different structured seed networks and grow them under two different rent structures and sizes. Subsequently, we will use the method mentioned ustering results. Our goal is to elucidate how the clustering results of the ıe networks and the structural characteristics of the nodes in the networks r different rules.

S

We build four different network growth processes to reveal the strong link between network clustering outcomes and changes in network structure. Special attention is paid to how the network structure evolves with different seed networks and growth rules. Before delving into structural changes in networks, we want to introduce two key factors: the definition of seed networks and how network growth rules guide network evolution. The growth process of the network is influenced by these two key conditions. The first is the initial structure of the seed network, which determines how the network growth rules are applied

guide the seed network's growth further shape the way the network's
rstanding these two critical factors, we can better explain the relationship
twork structure.

a clear differentiation of their structures, which is shown in Fig. 3.

consisting of core nodes and peripheral nodes, with a distinct centrality
re nodes have higher connectivity and importance in the network, while
ıave relatively lower connectivity and status. In contrast, the seed network B
es have randomly connected to the same extent. In network B, there are no
ıt position in the network topology. Such design ensures significant structural
ovides a clear starting point for the subsequent study of network growth and
ining 100 nodes, and provided corresponding diagrams, illustrating their

Seed-Network B

tworks. Seed network A has a Core–Periphery structure, with only a few nodes
degree. Seed network B is a random network. Each of its nodes has almost the
ed networks are different, several different structural changes will occur
ks of different structures, we can observe the relationship between the

key steps: introducing new nodes and attempting to establish connections
ill attempt to establish connections with nodes in the seed network according
ı structural change paths, in our study, we introduce two kinds of rules based
e BA rule and the ER rule.

the Barabasi–Albert (BA) model [2], an important method for generating scale-
nly connected to a few other nodes, while a very small number of core nodes
. The degree distribution of this network follows a power-law form, meaning
vast majority of nodes have relatively low degrees, exhibiting highly uneven
d nodes are more likely to connect to the core nodes in the network, i.e., those
onnecting to low-degree peripheral nodes. Specifically, the probability $P_{BA}(i)$
roportional to the degree $d(i)$ of node i. Its mathematical expression is as

$$P_{BA}(i) = \frac{d(i)}{\sum_{i=1}^{n} d(i)}, \quad (12)$$

where n is the total number of nodes in the network at the current time.

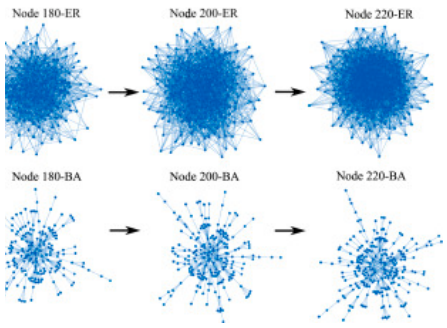
The second rule is the ER rule, originating from the Erdős–Rényi (ER) model [57]. In the ER model, network generation is
nerated randomly, and connections between nodes are independent, with the
nodes being equal.

(13)

etwork model that does not consider any specific structure or properties
of a new node connecting to existing nodes is equal and independent of the
lity of establishing a connection between any two nodes is fixed and
ing networks using the ER rule, we specify a connection probability of 0.1 for

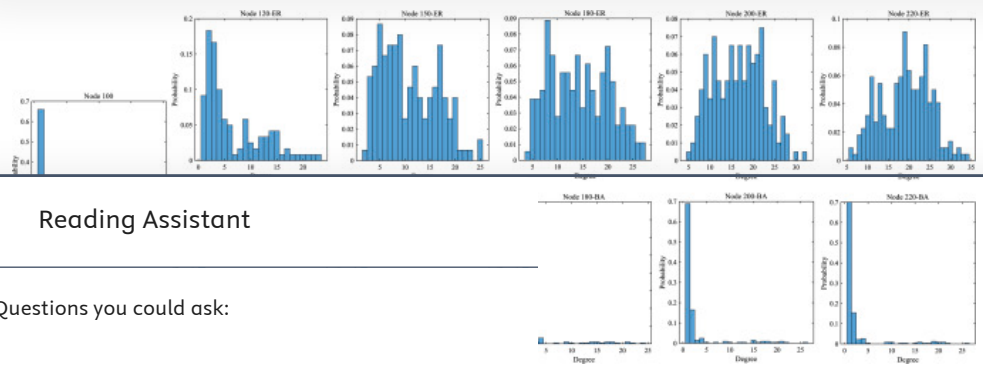
nder the guide of the BA rule to a network with 100 nodes. Obviously, it is a BA
l this result. Therefore, when Seed Network A continues to grow under
of the variations in the topological structure of Seed Network A under different
ture will be affected by the growth rules. The detailed topological structure
ıles are shown in Fig. 4.

ale, the network presents the characteristics of homogeneity. The connections
ıder the BA rule, newly added nodes are more inclined to connect to core
g in a small number of hub nodes with very high degrees. These hub nodes
while other nodes form the periphery of the network. Therefore, the network
ery structure under the BA rule. The change of the Seed-Network A's
er different rules can also be found in the change of degree distribution of each



l BA rules. subfigures in the top line show the change of networks' topological
ow the topological structure of change under the BA rule.

degree distribution under ER rules gradually changes from power law to
ower-law distribution, and a few nodes have higher degrees. With the growth
e random, resulting in the degree distribution gradually becoming uniform.
stribution always maintains the power law distribution. With the growth of
onnect to the existing highly connected nodes, which maintains the uneven



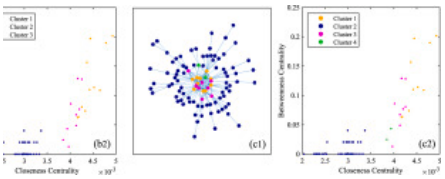
Reading Assistant

Questions you could ask:

Actions you could take:

: are grown from the Seed-Network A under different rules. The top line shows
wth is under the ER rule. The bottom line shows the degree distribution of the
ısly, Seed Network A is a BA network, and its power-law degree distribution is
les.

ork A and its performance under different growth rules, we conducted a
Fig. 6. The clustering results successfully partitioned the seed network into
distribution. In the subsequent chapters, we will continue to explore the



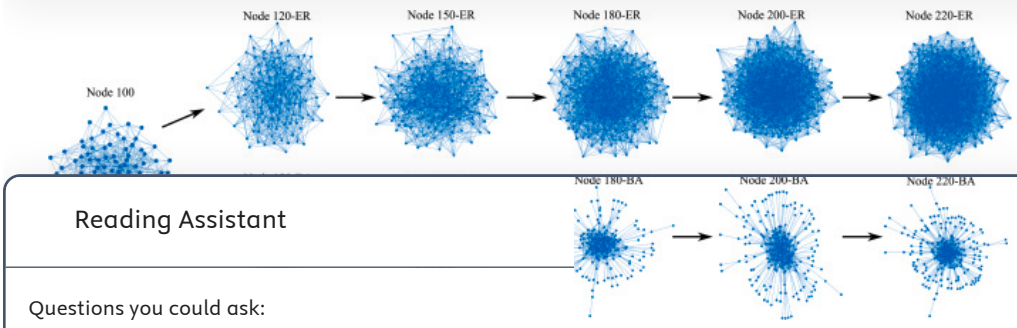
A. The clustering results of Seed Network A clearly delineate the core nodes
orm the network core, while scattered nodes constitute the periphery. The
re–periphery structure in Seed Network A.

wing from 1 node to 100 nodes guided by the ER rule. When the following
ıt rules, then the topological structural change will illustrate the relationship
s. The topological structure changes of Seed Network B under different rules

ule, there is no obvious core nodes. The probability of connection between
rk structure keeps the characteristics of uniform distribution. In contrast,
nnect to the hub node. Since the nodes in the seed network have similar
ıg with most nodes in the initial network is also equal, resulting in the initial
ws, it becomes easier for new nodes to connect to these core nodes, making the
ed structure to a core–periphery structure. There is a core community when
ructure changes are also illustrated in those networks' degree distribution as

Sign in to unlock the full response and ask your own questions.

Sign in



Reading Assistant

Questions you could ask:

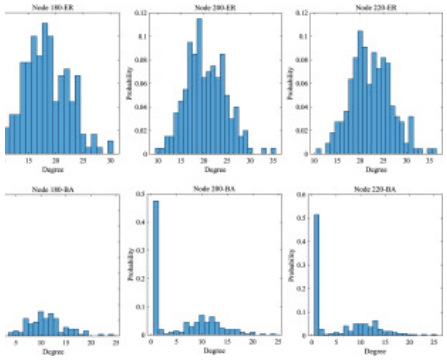
Actions you could take:

Sign in to unlock the full response and ask your own questions.

Sign in

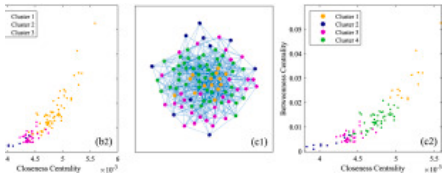
etwork B under ER and BA rules. There is a core community in the bottom

ıse of network size, the degree distribution of seed networks with uniform This shows that the connections of nodes show uniformity throughout the scale of the network increases, new nodes are more inclined to connect to aggravating the inhomogeneity of the network and the degree distribution of ıtion to power law distribution.



B's growth. Obviously, the is a special bulge in the degree distribution of

twork B, as shown in Fig. 9. The clustering results indicate that as the number ely uniformly distributed, reflecting its homogeneity characteristics. analysis during the growth process to gain a more comprehensive h rules influence the overall characteristics of the network.



ases, the node distribution in seed network B gradually becomes more ominent central nodes.

sis in the networks' growth

orks formed by different seed networks under various growth rules. We apply help us gain a deeper understanding of the characteristics of network ore the analysis, we will focus on the following aspects: Firstly, we will t the structural characteristics of the network. Secondly, we will explore

whether the changes in clustering results can reflect the differences between different network growth rules. By comparing clustering results under different growth rules, we will seek patterns influencing network structures and infer the evolutionary paths and mechanisms of networks under different growth rules. Finally, we will discuss whether clustering categories can indicate different node structural characteristics.

hange of the seed network A

ın the nodes' clustering in the process of the Seed Network A's growth under re based on the ER rule.

he network contains essential nodes in the core and edge nodes. When the show that the network is divided into core and peripheral nodes, which ıery seed network. With the increase in network size, the proportion of o matter what size the number of network nodes grows to, there will always e other categories have a relatively small percentage. This shows that with the ures of the core and edge nodes of the network grown by the BA rule become an clearly reflect the center-edge characteristics of the network. The change in vn in Table 2.

eed Network A is growing under the BA rule. The binary clustering has divided l the periphery nodes. They are labeled by class 1 and class 2. When the merge, but the biggest class is still class 2, which represents the periphery

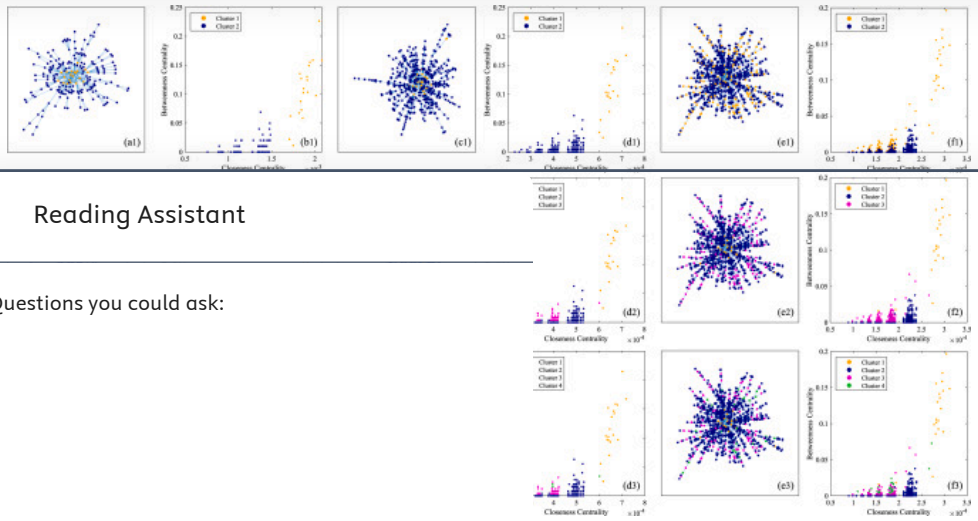
vork A.

	Node 3	Node 4	Node 5
	0.632	0.782	1.230
	0.051	0.100	0.176
	0.000	0.028	0.061
	0.028	0.000	0.061
	0.061	0.061	0.000

Results of Seed Network A under BA Rule.

cluster			Four-cluster			
	Class 2	Class 3	Class 1	Class 2	Class 3	Class 4
)	171 (85.5%)	9 (4.5%)	17 (8.5%)	171 (85.5%)	9 (4.5%)	3 (1.5%)
%)	438 (87.6%)	43 (8.6%)	19 (3.8%)	438 (87.6%)	39 (7.8%)	4 (0.8%)
%)	848 (84.8%)	131 (13.1%)	20 (2%)	848 (84.8%)	107 (10.7%)	25 (2.5%)

, new nodes tend to connect to existing, more tightly connected nodes, thus nodes. This becomes particularly evident when we divide the network into



Reading Assistant

Questions you could ask:

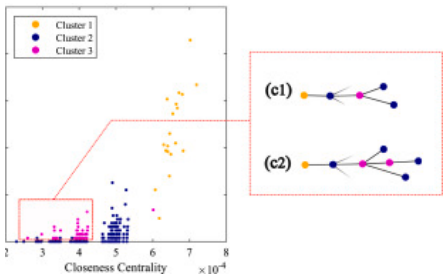
Actions you could take:

Sign in to unlock the full response and ask your own questions.

Sign in

k A under the BA model. In the Two-cluster, nodes with high closeness
ew class is the bridge nodes.

rst category of peripheral node clusters is located on the left side of the core
o the core nodes, thus possessing high closeness and betweenness centrality.
(a). In contrast, the second category of peripheral nodes is distributed at the
y connected to the core nodes but maintain indirect connections through the
nodes. Their betweenness centrality is relatively low, and the structural



es clustered into three categories and the structural characteristics of the node

iediary nodes, which play a crucial bridging role in the network by connecting
r network structures, the pink nodes are connected at one end to the first
o the second category of peripheral nodes, forming an important transitional
bridging function facilitates the flow of information between core nodes and
ectivity of the network, as illustrated in Fig. 11(c1).

complex, the pink intermediary nodes continue to play a bridging role,
heral nodes, as shown in Fig. 11(c2). Regardless of the complexity of the
onnectivity and stability by linking peripheral nodes of different hierarchies to
termediate nodes act as bridges between most peripheral nodes and core

r the ER rule, the structure changes, and the nodes' clustering is totally
the core–periphery network under the ER growth mechanism, each new node
vious nodes, resulting in a more evenly distributed structure. Through cluster
rule growth, the network structure gradually tends to be uniformly

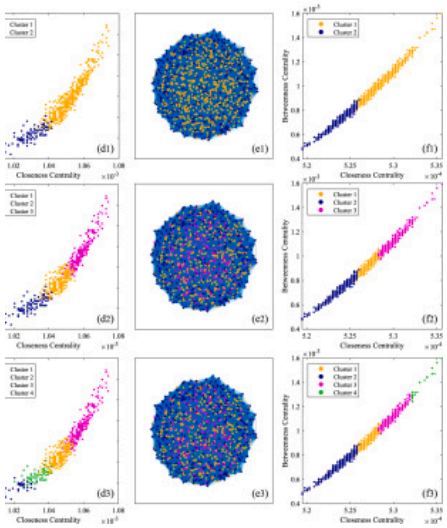
distributed, the existing core–edge structure gradually weakens, and the node connections show the characteristics of uniform distribution. This trend can be found in the percentage of nodes into different classes as shown in Table 3.

For networks with a core–periphery structure growing under the ER rule, we observe that the structural characteristics of the network and the performance of the scatter plot are very similar to those of random networks under the ER rule, as shown in Fig.

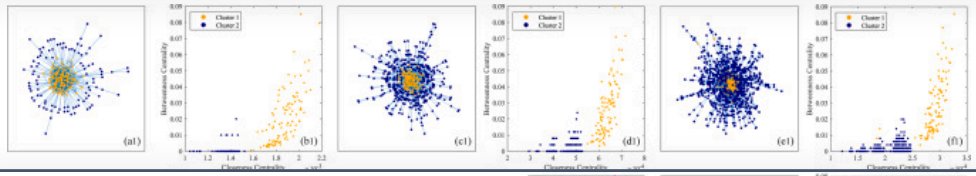
Results of Seed Network A under ER Rule.

ıster			Four-cluster			
	Class 2	Class 3	Class 1	Class 2	Class 3	Class 4
	56 (28%)	30 (15%)	114 (57%)	53 (26.5%)	30 (15%)	3 (1.5%)
	93 (18.6%)	182 (36.4%)	225 (45%)	37 (7.4%)	182 (36.4%)	56 (11.2%)
%)	380 (38%)	247 (24.7%)	373 (37.3%)	380 (38%)	227 (22.7%)	20 (2%)

categories of nodes are classified may vary, but clustering analysis still reveals ıub, and central nodes in the network. Specific explanations are provided as zes the dominant role of the ER rule in shaping network structure formation. R rule has a greater influence on networks' overall structure and organization, lexities exhibiting similar structural characteristics.

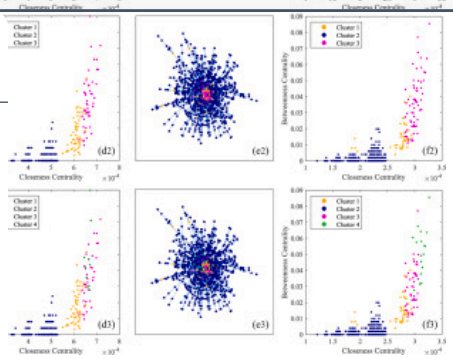


ring of Seed Network A under the ER model. The new classes come from the ıted uniformly.



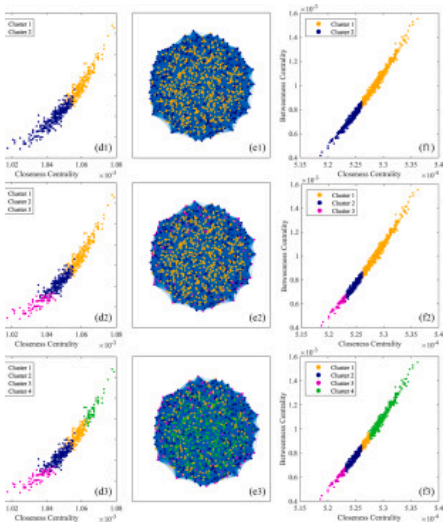
Reading Assistant

Questions you could ask:



of Seed Network B under the BA rule. The initial Seed Network B works as the
/th. The periphery nodes are always clustered into the same class. And the
:ore community.

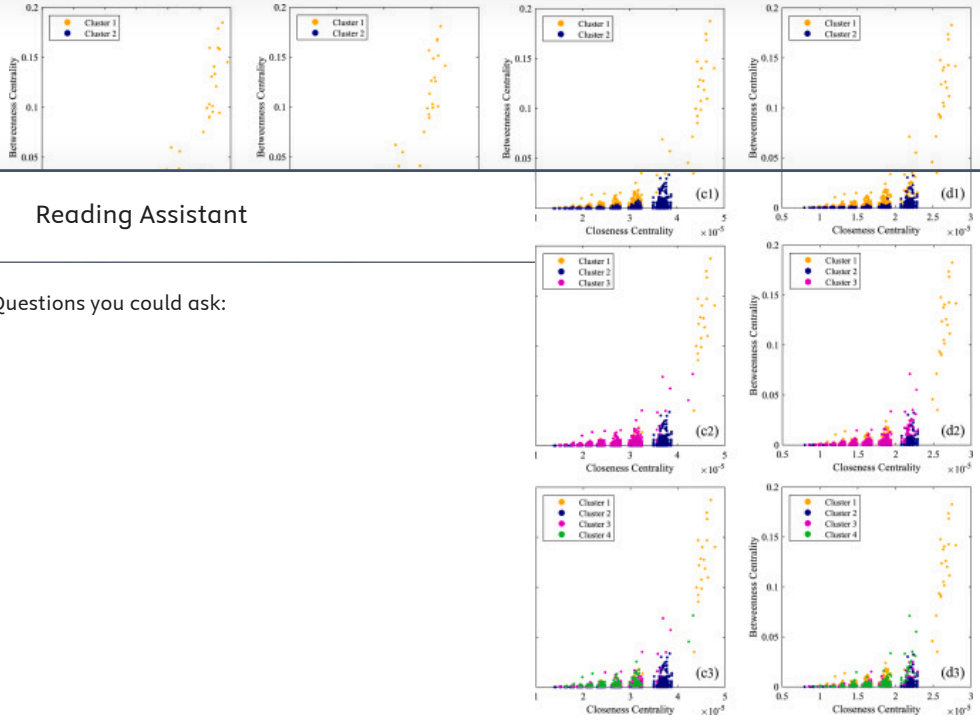
Actions you could take:



ıder the ER rule. Each class has the same size, and there is no special class in

Sign in to unlock the full response and ask your own questions.

Sign in

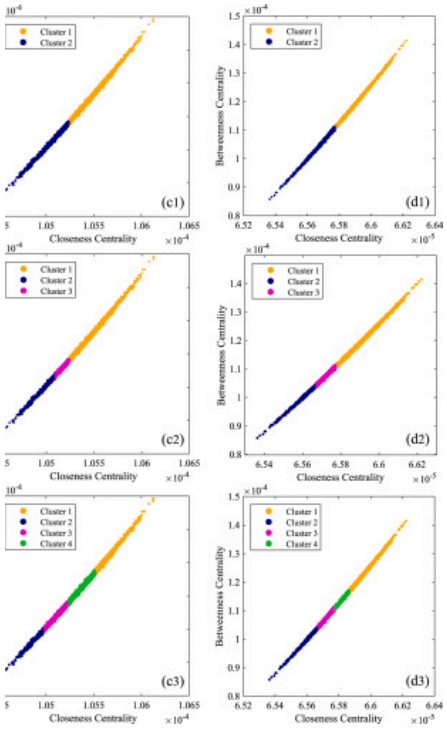


Reading Assistant

Questions you could ask:

Actions you could take:

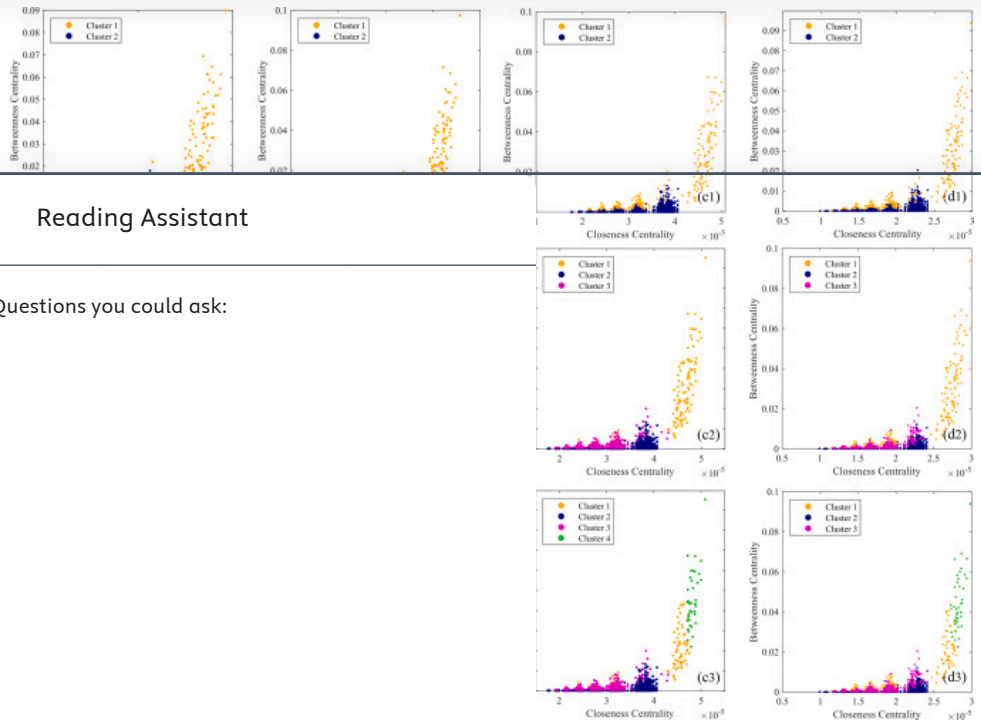
: are generated from a BA seed network under BA rule, with node number



: are generated from a BA seed network under ER rule, with node number

Sign in to unlock the full response and ask your own questions.

Sign in

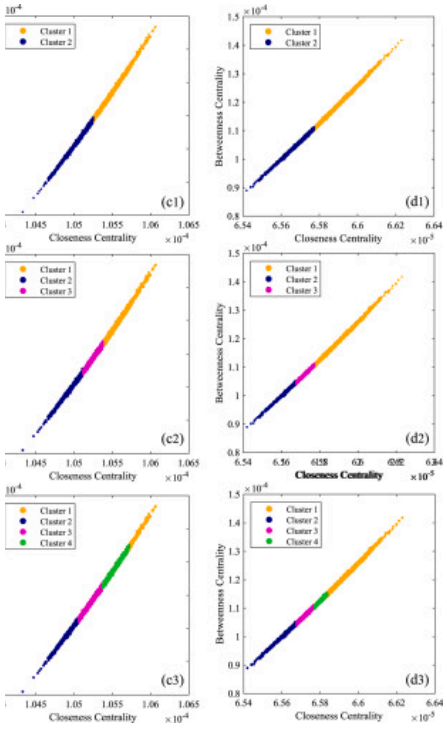


Reading Assistant

Questions you could ask:

Actions you could take:

te from ER seed networks under BA rule, with node number 2000, 3000, 5000



generated from a ER seed network under ER rule, with node number 2000,

Sign in to unlock the full response and ask your own questions.

Sign in

hange of the seed network B

les, and its degree distribution is a Poisson distribution. When it is under topological structure of those networks will be different, and so will the

When the Seed Network B is under the BA rule, the distribution of nodes in the initial stage is relatively uniform. following the increase of the network size, the probability of the occurrence of centralized nodes gradually increases. This is due to the growth mechanism under the BA rule, that is, newly added nodes are more likely to connect to nodes with higher degrees of existing nodes, resulting in a small number of nodes in the network with more connections, forming a centralized node. With the increase of network scale, node connections gradually show a trend of uneven distribution, and dominant node categories

vth rule, in which new nodes are more inclined to connect to nodes with
g the centrality of a few nodes in the network. The change of the number of
topological structure change in Seed Network B's growth, as shown in Table 4.

, when Seed Network B grows to 200 nodes, class 1 has 100 nodes, and class 2
ıave the same size. The initial seed network has 100 nodes, which means the
ıme class. Actually, when Seed Network B's size reaches node numbers 500
ɔf class 2 is increased. When the number of classes increases from 2 to 3 and 4,
ːter. It looks like there is a community in the network with 100 nodes, and the
ınity, i.e., there is a core in the network, and the core is the initial Seed

: Results of Seed Network B under BA Rule.

cluster			Four-cluster			
	Class 2	Class 3	Class 1	Class 2	Class 3	Class 4
%)	100 (50%)	42 (21%)	58 (29%)	100 (50%)	35 (17.5%)	7 (3.5%)
6%)	400 (80%)	47 (9.4%)	53 (10.6%)	400 (80%)	37 (7.4%)	10 (2%)
%)	896 (89.6%)	63 (6.3%)	41 (4.1%)	896 (89.6%)	46 (4.6%)	17 (1.7%)

ːs of the initial seed network and the BA model. Under the BA model, new
nodes. However, due to the uniform degree distribution of the nodes in the
connecting with nodes in the initial seed network is equal. This results in new
ıally forming new categories. Meanwhile, the nodes in the initial seed network
lters the structure of the core nodes but also complicates the relationships
ˢ among the core nodes continue to increase, further enhancing the

new categories are often subsets of the core community. This transition from
finement within the core reveals the evolutionary process of network structure
ɔre community play a crucial role in the network; they are centrally located,
ˈ, and have strong connections with other nodes, leading to efficient
ot only facilitates rapid information propagation but also effectively connects
exchange of information within the network, making it a key hub. The high
ıd of network centralization and the phenomenon of "the rich get richer". This
ling the structure of complex networks.

he growth of random seed networks growing under the ER rule is different
the size of the network increases, there will usually be one category of nodes
rule, we observe that with the gradual increase of cluster categories, there is
ate. Instead, we find that the proportion of node categories is relatively
t categories are not significant. Even in the case of large differences in the
ease of cluster categories, this part of nodes will be divided into new
ller. This shows that the random network formed under the ER rule has a

ıral features, the structural characteristics of the networks can be reflected. We
ed networks with core–periphery structures grown under the BA rule and
Ve examined networks with different numbers of nodes and the number of
ˈnary clustering, and quaternary clustering, which is shown in Table 5.

The first Two-cluster works on the network with 200 nodes. Class 1 has nearly 100 nodes, and class 2 also has nearly 100 nodes. There is no clear difference between the two classes. This phenomenon continues to network with 500 nodes in it. The two classes are still the same size. However, when the network's size reaches 1000 nodes, class 1 is bigger than class 2. The difference in the topological structure of those nodes emerges. This finding is also verified in Fig. 14.

: Results of Seed Network B under ER Rule.

	ıster		Four-cluster			
	Class 2	Class 3	Class 1	Class 2	Class 3	Class 4
)	74 (37%)	20 (10%)	74 (37%)	74 (37%)	20 (10%)	32 (16%)
%)	213 (42.6%)	68 (13.6%)	160 (32%)	213 (42.6%)	68 (13.6%)	59 (11.8%)
%)	296 (29.6%)	78 (7.8%)	233 (23.3%)	296 (29.6%)	78 (7.8%)	393 (39.3%)

rule in its process of growth; the whole process is equivalent to the classical ER t model has the same local structure, so this clustering of nodes based on the ɔde will be homogeneity.

node number 2000, 3000, 5000 and 8000. We can find that from left to right, ɔdes, nodes in the same cluster still located in the same position will the ıg of networks size have not show a distinguish difference on their node 15, Fig. 16, Fig. 17, Fig. 18).

:hod based on structural similarity, utilizing the local structure of the network ween nodes. We consider the structural characteristics of each node as a to measure the differences between these pieces of information, thereby s. Compared to traditional methods based on local structure, our method not ier to use than methods based on global structure. Our clustering method is improved the update rules for cluster centers, the distance calculation the initialization process, thus better capturing the intrinsic structure and

works with different initial configurations: core–periphery structure and e–periphery seed network features a structure with identifiable central and structure approximates real-world random networks. These seed networks es that newly introduced nodes are more likely to connect to the core nodes in h degrees, rather than to low-degree peripheral nodes. The ER rule indicates isting nodes is equal, regardless of the existing nodes' degrees. By growing s with different structures exhibiting significant differences in topology and is on these networks, and the results show that our clustering method the networks.

ıt network structures effectively reflect the structural characteristics of the of core–periphery structured networks grown under the BA rule and random se two structural networks into different categories and recorded the number s significant differences in the distribution characteristics of different types of eriphery network, when clustered into two categories, one category dominates nomenon becomes more pronounced as the network size and clustering ıe category with fewer nodes corresponds to core nodes, while the category This indicates the "rich get richer" phenomenon, where core nodes become size increases. Important nodes are in the minority, but they have high luence further strengthens. In contrast, the proportion of node categories in this feature remains unchanged as the network size increases. Furthermore, h patterns of the network. By comparing random networks under the BA rule e found that different growth rules have a significant impact on the network

structure. Under the BA rule, as the network size increases, the probability of central nodes appearing in a uniformly distributed network gradually increases. In contrast, in core–periphery structured networks grown under the ER rule, the network structure tends to be more uniform, with no obvious central nodes. This suggests that the structural characteristics of the network are mainly determined by its growth rules rather than the initial structure. All the results show us that the clustering of nodes in the networks based on the topological structure similarity is a reliable method that can be used to check the rules that guide the

es.

nt

, Formal analysis, Data curation. **Meizhu Li:** Writing – review & editing, . Qi Zhang: Writing – review & editing, Writing – original draft, Supervision, ı, Funding acquisition.

peting financial interests or personal relationships that could have appeared to

e Foundation of China (Grant No. 62303198), Research Initiation Fund for)1170008) and the Scientific Research Fund ing of Jiangsu University of Science

tworks

ar ↗

tructural correlations in weighted networks

ar ↗

el, Michael Cornell, Stephen G Oliver, Stanley Fields, Peer Bork
data sets of protein–protein interactions

ulio Superti-Furga
interdisciplinary signaling approaches

View in Scopus ↗ Google Scholar ↗

[7] Elad Schneidman, Michael J Berry, Ronen Segev, William Bialek

Weak pairwise correlations imply strongly correlated network states in a neural population

Nature, 440 (7087) (2006), pp. 1007-1012

Crossref ↗ View in Scopus ↗ Google Scholar ↗

View in Scopus ↗ Google Scholar ↗

[19] Réka Albert, Albert-László Barabási
Statistical mechanics of complex networks
Rev. Modern Phys., 74 (1) (2002), p. 47

View in Scopus ↗ Google Scholar ↗

tworks based on nonextensive statistical mechanics
1650118

ns for pattern identification

ar ↗

m permutation set
3
↗ Google Scholar ↗

tems with local constraints

ar ↗

J Watts
ks

Martin Chavez, D.-U. Hwang
mics

↗ Google Scholar ↗

Shlomo Havlin
networks

ar ↗

Li
es' betweenness centrality in complex networks
3
↗ Google Scholar ↗

nodes in network of networks

↗

l of complex networks
5
↗ Google Scholar ↗

, Fredrik Liljeros, Lev Muchnik, H Eugene Stanley, Hernán A Makse
n complex networks

ar ↗

[31] Qi Zhang, Meizhu Li, Yong Deng
Measure the structure similarity of nodes in complex networks based on relative entropy
Phys. A, 491 (2018), pp. 749-763
View PDF View article View in Scopus ↗ Google Scholar ↗

Reading Assistant

Questions you could ask:

Actions you could take:

Sign in to unlock the full response and ask your own questions.

Sign in

hen, Y.-Z. Ni
ers via local structural similarity
ar ↗

icture in networks

, Guohua Wu

ew in Scopus ↗ Google Scholar ↗

community structure in networks
ar ↗

ion
ar ↗

↗ Google Scholar ↗

niques
ces in Clustering, Springer (2006), pp. 25-71
ar ↗

ogical networks
6

Wang, Xiao Wang, Weixiong Zhang
ı with deep learning

icant communities and hierarchies, using message passing for
149
ar ↗

ies forecasting based on visibility graph
↗ Google Scholar ↗

[43] Sidney J Blatt, Charles A Sanislow, III, David C Zuroff, Paul A Pilkonis
Characteristics of effective therapists: further analyses of data from the national institute of mental health treatment of depression collaborative research program
J. Consult. Clin. Psychol., 64 (6) (1996), p. 1276
View in Scopus ↗ Google Scholar ↗

A new distance measure between two basic probability assignments based on penalty coefficient
Inform. Sci. (2024), Article 120883
View PDF View article View in Scopus ↗ Google Scholar ↗
[56] Jason Lu, Michael Small
A mutual information statistic for assessing state space partitions of dynamical systems

Reading Assistant

Questions you could ask:

ɪandom graph

ɔn the euclidean distance of structural characteristics

Actions you could take:

Radius Constrained Arc-Limited Penetrable Visibility Graph: Concepts

ɪnd data mining, AI training, and similar technologies.

Sign in to unlock the full response and ask your own questions.

Sign in