A local-world evolving hypernetwork model*

Yang Guang-Yong(杨光勇) and Liu Jian-Guo(刘建国)†

Research Center of Complex Systems Science, University of Shanghai for Science and Technology, Shanghai 200093, China

(Received 28 May 2013; revised manuscript received 30 July 2013; published 12 November 2013)

Complex hypernetworks are ubiquitous in the real system. It is very important to investigate the evolution mechanisms. In this paper, we present a local-world evolving hypernetwork model by taking into account the hyperedge growth and local-world hyperedge preferential attachment mechanisms. At each time step, a newly added hyperedge encircles a new coming node and a number of nodes from a randomly selected local world. The number of the selected nodes from the local world obeys the uniform distribution and its mean value is \( m \). The analytical and simulation results show that the hyperdegree approximately obeys the power-law form and the exponent of hyperdegree distribution is \( \gamma = 2 + 1/m \). Furthermore, we numerically investigate the node degree, hyperedge degree, clustering coefficient, as well as the average distance, and find that the hypernetwork model shares the scale-free and small-world properties, which shed some light for deeply understanding the evolution mechanism of the real systems.

Keywords: local-world evolving hypernetwork model, power-law form, small-world property

PACS: 89.75.Hc, 89.65.Ef, 05.70.Ln

DOI: 10.1088/1674-1056/23/1/018901

1. Introduction

The collaboration networks can be found everywhere in the real world, such as the actor-film network\,[1–3]\) papercientist network\,[4–8]\) and so on, which are usually described as simple networks.\,[9,10]\) Some literatures theoretically studied the degree distribution of collaboration networks based on Markov chain theory\,[11]\) and the rate equation approach.\,[12]\) In some cases, however, these models could only provide partial information of the real-life systems. For example, in the collaboration networks, we only know whether authors have collaborated or not, while we do not know how many papers an author has written, and we also do not know whether these authors coauthored the same paper or not. Bipartite network is a feasible solution to describe the collaboration system, where nodes are divided into two sets: the disjoint node set representing actors and the other one representing the collaborated papers.\,[13,14]\) Since there is no edge between any pairs of authors, the information between the authors is hard to map. For instance, we only know the fact that an author wrote how many papers and a paper was co-authored by how many authors, but it is hard to calculate the clustering character of these collaborated authors. As a new tool, the hypernetwork is a feasible way to model the collaboration systems.\,[15–17]\) In the hypernetworks, the edges, known as hyperedges, can contain more than two nodes.

Wang et al.\,[16]\) and Hu et al.\,[17]\) proposed a hypernetwork evolving model to mimic the collaboration evolution based on the growth and global preferential attachment mechanisms. However, in a scientific collaboration network, an author would look for coauthors in the similar research field and choose coauthors in his own local world. Therefore, the local preferential attachment mechanism was introduced to describe the evolution mechanism of the collaboration network.\,[18,19]\) Meanwhile, the number of nodes in the hyperedges should be different as the number of authors of different papers is always different.

Inspired by the above ideas, we present a local-world non-uniform evolving hypernetwork (LWH) model, which introduces the hyperedge growth and local-world preferential attachment mechanisms. Theoretical analysis and numerical simulation exhibit that the hyperdegree obeys the power-law form and the exponent of hyperdegree distribution is \( \gamma = 2 + 1/m \). Moreover, we numerically study the node degree, hyperedge degree, clustering coefficient, as well as the average distance and find that the LWH model shares the scale-free and small-world properties.

The rest of our paper is organized as follows. In Section 2, the LWH model with the growth and local-world preferential attachment mechanisms is given. In Section 3, we theoretically and numerically investigate the hyperdegree distribution and numerically study the other statistical properties of the LWH model, including the node degree, hyperedge degree, clustering coefficient, as well as the average distance. Conclusions and discussion are given in Section 4.

*Project supported by the National Natural Science Foundation of China (Grant Nos. 71071098, 91024026, and 71171136). Liu Jian-Guo is supported by the Shanghai Rising-Star Program, China (Grant No. 11QA1404500), and the Leading Academic Discipline Project of Shanghai City, China (Grant No. XTKX2012).

†Corresponding author. E-mail: liujg004@ustc.edu.cn

© 2014 Chinese Physical Society and IOP Publishing Ltd

http://iopscience.iop.org/cpb http://cpb.iphy.ac.cn

018901-1
2. The local-world hypernetwork model

Let $H = (V, E)$ be a simple and finite hypernetwork with the node set $V = \{v_1, v_2, \ldots, v_N\}$ and the hyperedge set $E = \{E_1, E_2, \ldots, E_l\}$, where $N$ is the node number, $I$ is the hyperedge number, and $E_j$ ($j = 1, 2, \ldots, l$) is a nonvoid sub-set of $V$. Let $r(H) = \max_j |E_j|$ and $s(H) = \min_j |E_j|$. If $r(H) = s(H)$, we say that $H$ is a uniform hypernetwork. Otherwise, $H$ is a non-uniform hypernetwork.\cite{20,21} Let $h = (V, e)$ be a simple and finite complex network generated by the evolving mechanism of the hypernetwork model with the node set $V = \{v_1, v_2, \ldots, v_N\}$ and edge set $e = \{e_1, e_2, \ldots, e_n\}$, where $n$ is the edge number.

For better understanding and describing the real-life collaboration system, the LWH model could be constructed in the following way (see Fig. 1).

(i) Initial condition The hypernetwork consists of $M_0$ nodes and $E_0$ hyperedges in the initial stage.

(ii) Determination of the local-world Select $M$ ($M < M_0$) nodes randomly from the existing hypernetwork as the local world at each time step.

(iii) Hyperedge growth Add a new hyperedge encircling a newly added node and $m_t$ selected nodes in the local world determined in step (ii) at time step $t$, where $m_t$ is a value selected randomly from the set $U = \{m - 1, m, m + 1\}$ and obeys a uniform distribution, and $m$ is a preset fixed value and $\frac{1}{2} \sum m_t = m$.

(iv) Local-world hyperedge preferential attachment Choose $m_t$ nodes in the local world to construct the new hyperedge. The probability $\prod$ for node $i$ is selected and depends on the hyperdegree $d_H(i)$ of node $i$, such that

$$\prod_i (d_H(i)) = \frac{M}{M_0 + t} \frac{d_H(i)}{\sum_{j \in L} d_H(j)}.$$  

where $L$ denotes the local world node set and the hyperdegree $d_H(i)$ is defined as the number of hyperedges node belonging to $i$. After $t$ time steps, this model leads to a hypernetwork with $(M_0 + t)$ nodes and $(E_0 + t)$ hyperedges. Figure 1 shows the evolving process of the LWH model at each time step.

3. Hypernetwork analysis

To investigate the effect of the local-world property and different mean values $m$ on the statistical characteristics of the LWH model, we explore two groups of experiments. The first one is to investigate the effect of the local-world size $M$ on the evolving results with the initial conditions of $M_0 = 160$, $E_0 = 158$, and $m = 2$. The second one is to explore the effect of the mean values $m$ on the evolving results. Initially, we set $M_0 = 160$, $E_0 = 158$, and $M = 150$. The size of local world $M$ is obtained according to the theory of Dunbar’s number that the cognitive limit number of social relationships we can maintain is around 150.\cite{22,23} As the size of the local world $M$ is 150 and the initial hypernetwork size $M_0$ should be larger than $M$, we let the initial condition $M_0 = 160$ and the choice of different $M_0$ will not affect the analysis and experiments. The statistical properties are investigated below.

3.1. Hyperdegree distribution

The hyperdegree $d_H(i)$ is defined as the number of the hyperedges node belonging to $i$. For instance, in scientific collaboration hypernetworks, the hyperdegree of node $i$ is the number of papers that the author $i$ wrote. The hyperdegree distribution $P(d_H)$ is defined as the probability that one randomly selected node possesses $d_H$ hyperedges, which is a very important quantity to characterize the hypernetwork structure. The theoretical results are given based on the mean-field approach.\cite{24}

3.1.1. Theoretical analysis

The generation process of the LWH model may be divided into two stages. The first stage generates an initial hypernetwork for local-world evolving process, which contains $M_0$ nodes and $E_0$ hyperedges. In the second stage, the hypernetwork model starts with $M_0$ nodes and $E_0$ hyperedges. At each time step $t$, a new hyperedge is added into the system and will encircle a new coming node and $m_t$ selected nodes in the local world. At time step $t$, the growing hypernetwork consists of $(M_0 + t)$ nodes enclosed by $(E_0 + t)$ hyperedges. When choosing $m_t$ nodes from the local world, the probability
that the node \(i\) is selected could be given by
\[
  m \left[ \prod_{L} d_{H}(i) \right] \left[ 1 - \prod_{L} d_{H}(i) \right]^{m-1} \approx m \prod_{L} d_{H}(i),
\]
where \(\frac{1}{m} \sum_{i} m_i = m, L, d_{H}(i)\) denote the local world node set and the hyperdegree of node \(i\), respectively. Let \(d_{H}(t)\) be the hyperdegree of node \(i\) at time \(t\). Consequently, \(d_{H}(t)\) satisfies the dynamical equation
\[
  \frac{\partial d_{H}}{\partial t} \approx m \prod_{(i)} d_{H}(i) = \frac{mM}{M_0 + t} \sum_{j=L} d_{H}(j). \tag{3}
\]
Because the random selection of the \(M\) nodes contributes to a local world connection at each time step \(t\), the cumulative degree of the local world depends on the random selection. \[18\]
In general, to simplify the following analysis, we assume that
\[
  \sum_{j=L} d_{H}(j) = \langle d_{H}(i) \rangle M, \tag{4}
\]
where the average hyperdegree
\[
  \langle d_{H}(i) \rangle = \frac{D_0 + (m + 1)}{(M_0 + t)},
\]
and \(D_0\) is the sum of all nodes’ hyperdegree in the initial hypernetwork generated by the first stage. Substituting Eq. (4) into Eq. (3) leads to
\[
  \frac{\partial d_{H}}{\partial t} = \frac{m}{D_0 + (m + 1)} d_{H}(t). \tag{5}
\]
When \(D_0\) is small and \(t\) is large, equation (5) is approximately equal to
\[
  \frac{\partial d_{H}}{\partial t} = \frac{m}{(m + 1)} d_{H}(t). \tag{6}
\]
Suppose that node \(i\) is added into the hypernetwork at time \(t_i\). Since the initial condition that every node \(i\) is \(d_{H}(t_i) = 1\), the solution of the above equation would be
\[
  d_{H}(t) = \left( \frac{t}{t_i} \right)^{m/m+1}. \tag{7}
\]
Using Eq. (7), the probability that a node has a hyperdegree \(d_{H}(t)\) smaller than \(d_{H}\), \(P(d_{H}(t) < d_{H})\), can be written as
\[
  P(d_{H}(t) < d_{H}) = P\left( t_i > \frac{t}{d_{H}^{(m+1)/m}} \right). \tag{8}
\]
Suppose the nodes are added equally into the hypernetwork at each time step, the \(t_i\) values have a constant probability density \(p(t_i) = 1/t\). Substituting it into Eq. (8) yields
\[
  P(t_i > \frac{t}{d_{H}^{(m+1)/m}}) = 1 - \frac{t}{d_{H}^{(m+1)/m}}. \tag{9}
\]
The hyperdegree distribution \(P_{H}(d_{H})\) can be obtained by
\[
  P(d_{H}) = \frac{\partial P(d_{H}(t) < d_{H})}{\partial d_{H}} = \frac{(m+1)}{m} d_{H}^{(2+1/m)}. \tag{10}
\]
Thus, the probability distribution \(P(d_{H})\) has a generalized power-law form
\[
  P(d_{H}) = \frac{m+1}{m} d_{H}^{-(2+1/m)}, \tag{11}
\]
where the power exponent equals \(\gamma = (2 + 1/m)\).

### 3.1.2. Numerical simulations

The numerical simulations of the hyperdegree distribution \(P(d_{H})\) is given in Fig. 2. Figure 2(a) shows the hyperdegree distribution \(P(d_{H})\) with the mean value \(m = 2\) and different local-world sizes \(M\). Figure 2(b) shows the case of the same local-world size \((M = 150)\) and different mean values \(m\). As seen from Fig. 2(a), the hyperdegree distribution \(P(d_{H})\) of LWH model closely overlap together as \(M\) increases. Figure 2(b) shows that the power-law exponent \(\gamma\) of degree distribution decreases as \(m\) increases.

The simulation results are quite consistent with the theoretical ones. We can easily find that the hyperdegree distribution is approximately independent of the local-world size \(M\) and exhibits a pow-law distribution, i.e., \(P(d_{H}) \sim d_{H}^{-\gamma}\), where the exponent \(\gamma\) is correlated with the mean value \(m\) \((\gamma = 2 + 1/m, \gamma \in (2, 3)]\), and the power exponent of the empirical results of the social tagging hypernetwork, 2.28 and 2.13, is also in \((2, 3)]\). \[25\]

### 3.2. Degree distribution

The degree \(k\) of a node is the number of edges incident with the node. The node degree distribution \(P(k)\) is defined as the probability that a node is connected to \(k\) other nodes. We numerically investigate the degree distribution \(P(k)\) in Fig. 2. Figure 2(c) describes \(P(k)\) in the case of the same mean value \((m = 2)\) and different local-world sizes \(M\), and figure 2(d) shows \(P(k)\) in the case of the same local-world size \((M = 150)\) and different mean values \(m\), from which one can find that the degree distribution is also approximately independent of the local-world size \(M\) and exhibits a pow-law distribution, i.e., \(P(k) \sim k^{-\delta}\), where \(\delta\) is correlated with the mean value \(m\). Since all nodes in a hyperedge are fully connected, the more hyperedges a node is attached to, the more other nodes the node connects to. Thus, when a new hyperedge is added according to the hyperedge preferential attachment mechanism, a node that acquires more connections than others will increase its selected probability, which is similar to the growth mechanism of Barabási–Albert (BA) model. \[27\]
3.3. Hyperedge degree distribution

The hyperedge degree $d_{hd}$ is defined as the number of hyperedges one hyperedge connects to. For example, in scientific collaboration hypernetworks, the $d_{hd}(i)$ denotes the number of papers that are coauthored by the authors of paper $i$. In the projected hypernetwork, the hyperedge degree corresponds to the node degree (see Fig. 3).

Figure 4 shows the numerical simulations of the hyperedge degree distribution $P(d_{hd})$. Figure 4(a) shows $P(d_{hd})$ in the case of the same mean value ($m = 2$) with different local-world sizes $M$, and figure 4(b) presents $P(d_{hd})$ in the case of the same local-world size ($M = 150$) with different mean values $m$. At each time step, a newly added hyperedge encircles a number of nodes in the local world, so that the hyperedge degree of the hyperedges belonging to the selected nodes will increase. From Fig. 4(a), we observe that for small values of $d_{hd}$, the hyperedge degree distribution deviates from the power-law behavior, which originates from the effect of the fluctuation in the number of new nodes in a new hyperedge. While, for big values of $d_{hd}$, it displays a power-law behavior and the exponent is approximately equal to 5, which manifests that the hyperedges show a trend of preferential attachment due to increased scale of the hypernetwork. From Fig. 4(b), one can find that for small values of $d_{hd}$, as the value $m$ increases, the number of hyperedges with the same $d_{hd}$ is less. This is mainly because the more nodes a hyperedge encircles, the more hyperedges the hyperedge connects to. Consequently, on the whole, each hyperedge connects to more other hyperedges, which results in less hyperedges with small values of $d_{hd}$.
3.4. Clustering coefficient

Clustering, also known as transitivity, is a typical property of acquaintance networks, where two individuals with a common friend are likely to know each other. The average clustering coefficient $\langle C \rangle$ quantifies the extent to which nodes adjacent to a given node are linked. Let $E_i$ denote the set of neighbor nodes of node $i$ with degree $k_i$ and $|E_i|$ denote the number of edges between the neighbor nodes, then local clustering coefficient $C_i$ of node $i$ and $\langle C \rangle$ are, respectively, defined as follows:

$$C_i = \frac{|E_i|}{k_i(k_i - 1)/2},$$

$$\langle C \rangle = \frac{1}{N} \sum_i C_i. \quad (13)$$

In the hypernetworks, the average clustering coefficient is defined as the degree of overlap between different hyperedges that a hyperedge participates in. In the projected hypernetwork, the hyperedges can be connected by a common node (see Fig. 4). Thus, we calculate the average clustering coefficient $\langle C_{\text{H}} \rangle$ using Eq. (13) in the projected hypernetwork as follows:

$$\langle C_{\text{H}} \rangle = \frac{1}{N_{\text{E}}} \sum_i C_{i,\text{H}}, \quad (14)$$

where $N_{\text{E}}$ indicates the node number in the projected hypernetwork and $C_{i,\text{H}}$ represents the clustering coefficient of node $i$ in the projected hypernetwork.

Figures 5(a) and 5(e) show that both the complex network and the projected network exhibit a high clustering degree in three cases of $M = 15, 50$, and $150$. One can find that $\langle C_h \rangle$ and $\langle C_{\text{H}} \rangle$ are both bigger than 0.45 and approximately remain unchanged as the network size $N$ and the hyperedge size $N_{\text{E}}$ increases. The bigger the value $M$, the higher the values $C_h$ and $C_{\text{H}}$. This is mainly because the bigger the value $M$, the greater the probability that nodes with higher clustering will be selected. Meanwhile, both the complex network (Fig. 5(b)) and the projected network (Fig. 5(f)) also exhibit a high clustering degree in three cases of $m = 2, 4, \text{and } 6$. One can see that both $\langle C_h \rangle$ and $\langle C_{\text{H}} \rangle$ approximately remain unchanged as the network size $N$ and the hyperedge size $N_{\text{E}}$ increase, respectively, which are both bigger than 0.35. In Fig. 5(b), the bigger the value $m$, the higher the value $\langle C_h \rangle$. This is mainly because the bigger the value $m$, the more the coupled links between the new nodes at each time step. While in Fig. 5(f), the bigger the value $m$, the lower the value $\langle C_{\text{H}} \rangle$, which is different from Fig. 5(b). This may be mainly because the bigger the value $m$, the more unclear the preferential mechanism.

3.5. Average distance

Another important quantity is the average distance between a random pair nodes in a network. Hence, the average distance, $\langle D \rangle$, which is defined as the average value of the shortest paths between any two nodes in the network, is used to measure the efficiency of retrieving a target node in a network. Let $d_{ij}$ denote the length of shortest path between node $i$ and node $j$, then

$$\langle D \rangle = \frac{\sum_{i \neq j} d_{ij}}{N(N-1)}, \quad (15)$$

where $N$ is the node number of the network. Since $\langle D \rangle$ characterizes the ability of two nodes to communicate with each other, it is often used to describe the interconnectedness of the network.

In the hypernetwork, the hyperedge–hyperedge distance is defined as the shortest paths between two hyperedges that are reachable along joint nodes. In the projected hypernetwork, the hyperedge–hyperedge distance represents the shortest paths between two nodes via the edges. Hence, the average distance of hypernetworks, $\langle D_{\text{H}} \rangle$, is defined as the average value of all hyperedge–hyperedge distances.

We investigate the effect of the hypernetwork size on the average distance $\langle D_h \rangle$ between nodes in the complex network and the average distance $\langle D_{\text{H}} \rangle$ between hyperedges in the projected hypernetwork. From Figs. 5(c) and 5(d), the average distances $\langle D_h \rangle$ increase very slowly as $N$ increases, which are smaller than 4.5 and become smaller as $M$ and $m$ increase. From Figs. 5(g) and 5(h), the average distances $\langle D_{\text{H}} \rangle$ also increase very slowly as $N_{\text{E}}$ increases, which are also smaller than 4.5 and become smaller as $M$ and $m$ increase. In all cases, the simulation results suggest that both the complex network and the hypernetwork exhibit the small-world property.
4. Conclusions and discussion

4.1. Summary

Our proposed hypernetwork model and the bipartite network model [29] both contain more information than the collaboration network model. [30] However, for the bipartite network model [29] the “homogeneity” in the definition of nodes is lost for certain nodes representing actors and others representing acts of collaboration. Since no edge connects the nodes in the same property set, the information between the same nodes is hard to map. Taking into account the fact that one author often looks for collaborators from his local world [31–33] and the number of collaborators is always different, we propose the LWH model by combining the hyperedge growth and local-world hyperedge preferential attachment mechanisms to investigate effects of the local-world connectivity and different mean values m on the statistical properties of the hypernetwork structure.

We investigate the hyperdegree distribution by theoretical analysis and numerical simulation. The simulation results agree with the analytic ones very well, indicating that the hyperdegree distribution conforms to a power-law distribution with the exponent \( \gamma = 2 + 1/m \), where \( m \) is the mean value of nodes selected from the local world. We also numerically study the other statistical properties of the LHW model, including the node degree, hyperedge degree, clustering coefficient, as well as the average distance. In our model, the hyperedge is added in terms of the local world, but not the global network information. Both the hyperdegree distribution and the degree distribution are approximately independent of the local-world size \( M \) and exhibit the power-law form, where the exponent depends on the mean value \( m \). Meanwhile, this model exhibits the highly clustering and short average distance. In other words, this model shares the scale-free and small-world properties.

4.2. Limitations and future work

In this paper, the local-world preferential attachment mechanism is adopted to investigate the effect of the local-world connectivity and different mean values \( m \) on the statistical properties of the hypernetwork structure, and interesting results have been obtained. However, more elaborate endeavors are needed. A few research issues that directly follow this work can be briefly discussed.

First, in this work, we only consider the situation that all authors of one paper only coauthor a paper together. In the real life, however, all authors of one paper may collaborate more than once together. In the future work, we will investigate the hyperedge weighted hypernetworks by introducing hyperedge weight.

The second limitation of the present work is that the creation of new hyperedges constructed by the old nodes and the
lifetime of nodes are neglected in our model. However, in the real collaboration system, the old nodes always collaborate with each other. Furthermore, in the scientific collaboration hypernetwork and movie actor collaboration hypernetwork, the lifetime of nodes is finite. Therefore, as another direction of the future work, we need to revise our model to cater for such situation. When the above situation is taken into account, there will be more prominent influence on the overall statistical properties in the hypernetwork. If we take the lifetime of nodes into account in a hypernetwork, the question of whether the hyperdegree distribution still follows the power-law form remains in debate. It can be anticipated to obtain some new interesting findings in this research direction.

Third, in this work there are assumptions that nodes are added to the hypernetwork at equal time intervals, and the arrival time of nodes follows the uniform distribution. Actually, in more cases, the growth of the nodes are random and the newly added nodes may perform nonlinear preferential attachments. In the future, it is a meaningful work to analyze growing hypernetworks without the assumption of continuum and the existence of the stationary degree distribution.

Last but not least, we investigate the statistical properties of the LWH model, but the corresponding empirical studies are still absent. As a complement to this work, we model the real cooperation cases and statistically analyze the topological characteristics. Due to the intrinsic intractability of the cooperative behavior, such empirical study is essentially challenging, but this issue is worthy of inquiry in the context of social collaboration.

References
[18] Li X and Chen G R 2003 Physica A 328 274
[27] Barabási A L and Albert R 1999 Science 286 509
[33] Ren F X, Shen H W and Cheng X Q 2012 Physica A 391 3533
[34] Wang L N, Guo J L, Yang H X and Zhou T 2009 Physica A 388 1713