Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Computer Science 9 (2012) 1400 – 1409

Procedia
Computer Science

International Conference on Computational Science, ICCS 2012

A Social Network Model Exhibiting Tunable Overlapping Community Structure

Dajie Liu[a,*], Norbert Blenn[a], Piet Van Mieghem[a]

[a] *Faculty of Electrical Engineering, Mathematics and Computer Science*
Delft University of Technology, P.O. Box 5031, 2600 GA Delft, The Netherlands

Abstract

Social networks, as well as many other real-world networks, exhibit overlapping community structure. In this paper, we present formulas which facilitate the computation for characterizing the overlapping community structure of networks. A hypergraph representation of networks with overlapping community structure is introduced. Using the hypergraph representation we setup a social network model which exhibits innate tunable overlapping community structure. By comparing simulation results of our model with results of the Hyves network and the SourceForge network, we show that our model captures most of the common features of large social networks. We analytically give the relation between the maximum membership number of a network and the adjacency eigenvalues of the simple graph representation of the network, which is the line graph of the hypergraph representation.

Keywords: social networks, overlapping community structure, preferential attachment, hypergraph

1. Introduction

Social networks, as one type of real-world complex networks, are currently widely studied [1, 2, 3]. Most social networks possess common properties of the real-world networks, such as high clustering coefficient, short characteristic path length and power law degree distribution [4, 5]. Particularly, they possess some special properties like assortative mixture, community and hierarchical structure [3, 6, 7, 8]. The communities are the subnets, which exhibit relatively higher levels of internal connections. Community structures feature important topological properties that have catalyzed researches on community detection algorithms and on modularity analysis [9, 10, 11]. The communities overlap with each other when nodes belong to multiple communities. The overlap of different communities exists widely in real-world complex networks, particularly in social and biological networks [12, 13, 14]. Human beings have multiple roles in the society, and these roles make people members of multiple communities at the same time, such as companies, universities, families or relationships, hobby clubs, etc. Proteins may also involve in multiple functional categories in biological networks, which explains why and how overlapping communities emerge in social and biological networks. In the movie actor network, where nodes are the actors and two actors are connected if they have been casted together in one or more movie, we could regard the set of actors in one movie as a community,

[*] Corresponding author
Email address: d.liu@tudelft.nl (Dajie Liu)

doi:10.1016/j.procs.2012.04.154

and this community is a fully connected subnetwork which are called cliques or complete subgraphs in the language of graph theory. According to the definition of movie actor network, the communities of all the movies are cliques. These communities overlap with each other if they have actors in common. The similar networks are the science coauthorship networks (nodes represent the scientists and two nodes are connected if they have coauthored one or more articles and the articles are communities), the journal editor networks (nodes as the editors and two editors are adjacent if they serve on the same editorial boards of journals) and sports player networks (nodes as players and two players who played in the same games are connected). These social networks naturally contain many cliques.

Palla et al. [12] defined four metrics to describe how the communities of networks overlap with each other: the membership number of an individual, the overlapping depth of two communities, the community degree and the community size. Palla et al. [12] showed that the communities of real-world networks overlap with each other significantly. They reported that the membership number of an individual and the overlapping depth of two communities and the community size all follow a power law distribution, except that the community degree features a peculiar distribution that consists of two distinct parts: an exponential distribution in the beginning and a power law tail. Pollner et al. [15] proposed a toy model of which both the community size and the community degree follow a power law distribution, by applying preferential attachment to community growth. There have been many efforts devoted to the modeling of social networks [16, 17, 18]. The growing networking model proposed by Toivonen et al. [18] succeeds in reproducing the common characteristics of social networks: community structure, high clustering coefficient and positive assortativity. The degree distribution of this model is somewhat deviating from a power law distribution despite being heavy-tailed.

We propose a complete set of metrics which can fully characterize the overlapping community structure of networks. We represent social networks by hypergraphs. The hypergraph representation of networks facilitates of the computations of the characterizing metrics. We establish a hypergraph-based social network model which exhibits innate tunable overlapping community structure. By comparing simulation results of our model with results of real-world networks, we show that our hypergraph model exhibits the common properties of large social networks: the community size, the community degree and the community overlapping depth all follow a power law distribution, and our model possesses high clustering coefficient, positive assortativity, short average path length. By tuning the input individual membership number to follow a power law distribution, the individual degree and the interest-sharing number also follow a power law distribution. We prove that all eigenvalues of the adjacency matrix of the line graph of the hypergraph, which represents a network, are not smaller $-m_{\max}$, where $m_{\max}$ is the maximum membership number of that network.

2. Definition of metrics for characterizing the overlapping community structure

In social networks, each individual can be characterized by the degree and the number of communities to which the individual belongs. A pair of individuals can be characterized by the number of common communities, which indicates how many common interests they share. Similarly, the size of a community, which is the number of individuals it contains, and the degree of a community, which is the number of other communities with whom individual(s) are shared, give additional information of a community. The number of individuals that two communities share can suggest how much they overlap with each other. All the six metrics for characterizing the overlapping community structure are defined formally as follows:

Definition 1. *(a) The degree d_j of an individual j in a network is defined by the number of individuals that connect to j; (b) The membership number m_j of an individual j is defined by the number of communities of which j is a member. The membership number m_j, together with the degree d_j, reflect the social connection of individual j; (c) The interest-sharing number $\alpha_{i,j}$ of individual i and j is the number of communities to which they both belong; (d) The community size s_j of community j equals the number of nodes that belong to community j; (e) The community degree u_j of community j is the number of other communities sharing individual(s) with community j; (f) The overlapping depth $\beta_{i,j}$ of two communities i and j equals the number of individuals that they share.*

Palla et al. [12] have defined the individual membership number m_j, the community size s_j, the community degree u_j and the community overlapping depth $\beta_{i,j}$. This paper augments the set of metrics to characterize the overlapping community structure with predefined individual degree d_j and the interest-sharing number $\alpha_{i,j}$. The

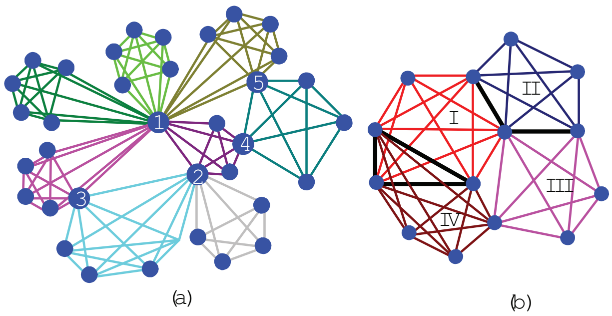


Figure 1: The example graph to illustrate the community structure. The nodes denote individuals. The communities consist of links of the same color and the shared thick black link(s), and the nodes incident to the links of both colors.

Table 1: The names and the members of all the communities of the exemplary social network of NAS.

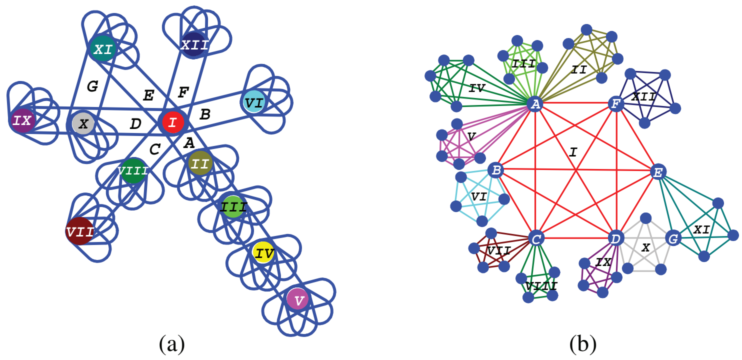
Index	Names of communities	Members (individuals)
I	NAS-TU Delft	A, B, C, D, E, F
II	A research group-MIT	$A, A_1, \cdots, A_5$
III	A research group-Cornell Univ.	$A, A_6, \cdots, A_{10}$
IV	IEEE/ACM ToN editorial board	$A, A_{11}, \cdots, A_{15}$
V	A research group-KSU	$A, A_{16}, \cdots, A_{20}$
VI	A research group-Ericsson	$B, B_1, \cdots, B_4$
VII	A research group-KPN	$C, C_1, \cdots, C_4$
$VIII$	Piano club	$C, C_5, \cdots, C_8$
IX	A research group-TNO	$D, D_1, \cdots, D_4$
X	A rock band	D, D_5, D_6, D_7
XI	A soccer team	$E, E_1, \cdots, E_4$
XII	Bioinformatics-TU Delft	$F, F_1, \cdots, F_4$

probability distributions of d_j, m_j, $\alpha_{i,j}$, s_j, u_j and $\beta_{i,j}$ play an important role in characterizing the community structure of a network.

We use the graphs in Figure 1 to exemplify the definitions of those metrics. The graph in Figure 1 (a) has labeled five nodes which are members of at least two communities. Using the metrics introduced in Definition 1, the graph illustrated in Figure 1 (a) has the following specifications: $d_1 = 24$, $d_2 = 12$, $d_3 = d_4 = 8$ and $d_5 = 9$. Nodes 1 – 5 belong to 5, 3, 2, 2 and 2 communities respectively, thus $m_1 = 5$, $m_2 = 3$ and $m_3 = m_4 = m_5 = 2$. Individual 1 and 2 belong to only one common community, hence $\alpha_{1,2} = 1$. As shown in Figure 1 (b), the communities $I - IV$ have 6, 5, 5 and 6 nodes respectively, hence, $s_I = s_{IV} = 6$ and $s_{II} = s_{III} = 5$. By the definitions, we have the overlapping width $\beta_{I,II} = 2$, $\beta_{I,III} = 1$, $\beta_{I,IV} = 3$, $\beta_{II,III} = 2$, $\beta_{II,IV} = 0$ and $\beta_{III,IV} = 1$, and the community degree $u_I = u_{III} = 3$, $u_{II} = u_{IV} = 2$.

3. The hypergraph representation of social networks

A hypergraph is the generalization of a simple graph, which is an unweighted, undirected graph containing no self-loops nor multiple links between the same pair of nodes. A hypergraph $H(N, L)$ has N number of nodes and L number of hyperlinks. Its nodes are of the same type as those of a simple graph, as shown in Figure 2 (a). However its hyperlinks can connect multiple nodes, like hyperlink A in Figure 2 (a) connecting nodes $I, II, \cdots, V$. The hyperlinks of a hypergraph should not be confused with hyperlinks on a webpage. A hypergraph is linear if each pair of hyperlinks intersects in at most one node. Hypergraphs where all hyperlinks connect the same number m of nodes are defined as m-uniform hypergraphs with the special case that 2-uniform hypergraphs are simple graphs.



(a) (b)

Figure 2: (a) The hypergraph representation of the network described in Table 1. The hyperlinks are the blue ellipse-like closed curves. The nodes are the disks with different colors marked with Roman numerals. A node and a hyperlink are incident if the node is surrounded by the hyperlink. The hyperlinks and nodes represent the individuals and the communities respectively. Individuals participate in multiple communities, implying that the communites overlap with each other. (b) The line graph of the hypergraph in (a), which is a simple graph. The nodes here denote the individuals while the communities consist of links of the same color and the nodes which are incident to them. Note that this graph is also the line graph of the hypergraph.

Definition 2. *The line graph of a hypergraph $H(N, L)$ is defined as the graph $l(H)$, of which the node set is the set of the hyperlinks of the hypergraph and two nodes are connected by a link of weight t, when the corresponding hyperlinks share t node(s).*

3.1. An illustrative example

In this subsection, we introduce the representation of social networks by hypergraphs. Traditionally, unweighted and undirected networks are represented by simple graphs. We show that social networks with fully connected communities can also be represented by hypergraphs with nodes denoting the fully connected communities and the hyperlinks denoting the individuals. We give an exemplary social network and then represent it by a hypergraph. Table 1 describes a small social network based on the friend-colleagueship of members of the NAS research group (Network Architectures and Services Group at Delft University of Technology). Individuals A, B, C, D, E, F are members of NAS and the other individuals are the members of communities which overlap with the NAS community. Two individuals have friend-colleagueship if they are in the same community, hence all the communities in this exemplary social network are fully connected.

We represent this network by the hypergraph shown in Figure 2 (a). The nodes of the hypergraph denote the communities and the individuals are denoted by the hyperlinks. There are 12 communities as described in Table 1, corresponding to the 12 nodes of the hypergraph in Figure 2 (a), and there are 53 individuals among whom 6 NAS members with the membership number $m_A = 5$, $m_C = m_D = 3$, $m_B = m_E = m_F = 2$. If an individual belongs to multiple communities, the corresponding nodes are connected by the hyperlink specifying that individual.

Figure 2 (b) depicts the line graph of the hypergraph in Figure 2 (a), which represents the exemplary social network. In the line graph, the individuals are denoted by nodes and the communities are denoted by links of the same color and the nodes which are incident to those links. In spite of its small size, the communities and the overlap of communities are shown clearly. By definition, the line graph in Figure 2 (b) of the hypergraph in Figure 2 (a) is a simple graph.

3.2. Expressions of metrics for overlapping communities

The line graph $l(H)$ is a simple graph when $H(N, L)$ is linear, and the line graph $l(H)$ of a nonlinear hypergraph $H(N, L)$ is a weighted graph. The adjacency matrix $A^{l(H)}_{L\times L}$ of the line graph $l(H)$ of a hypergraph $H(N, L)$ can be computed from the unsigned incidence matrices $R_{N\times L}$ of hypergraphs

$$A^{l(H)}_{L\times L} = \left(R^T R\right)_{L\times L} - diag\left(R^T R\right) \quad (1)$$

where the entry r_{ij} of R is 1 if node i and hyperlink j are incident, otherwise $r_{ij} = 0$. Basically, the adjacency matrix $A^{l(H)}$ equals the matrix $R^T R$ setting all the diagonal entries to zero. The interest-sharing number $\alpha_{i,j}$ of individual i and j equals the entry $a_{ij}^{l(H)}$of $A^{l(H)}$

$$\alpha_{i,j} = a_{ij}^{l(H)} \tag{2}$$

The membership number m_j of an individual j equals,

$$m_j = \sum_{i=1}^{N} r_{ij} \tag{3}$$

The community size s_i of community i is

$$s_i = \sum_{j=1}^{L} r_{ij} \tag{4}$$

Let $W_{N\times N} = \left(RR^T\right)_{N\times N} - diag(RR^T)$, then the overlapping depth $\beta_{i,j}$ of two communities i and j equals,

$$\beta_{i,j} = w_{ij} \tag{5}$$

where w_{ij} is an entry of $W_{N\times N}$.

If $H(N, L)$ is linear, the individual degree $d_j = \sum_{i=1}^{L} a_{ij}^{l(H)}$. If $H(N, L)$ is nonlinear, d_j equals the number of nonzero entries in the jth row/column of $A^{l(H)}$. Similarly, the community degree u_j equals the number of nonzero entries in row/column j of W.

4. Model Description

As a common property, the individual degree of social networks follows a power law distribution [1][4]. Palla et al. [12] and Pollner et al. [15] reported that the community size also follows a power law distribution. Nacher et al. [19] and Manka-Krason et al. [20] showed that the nodal degree of line graphs of simple graphs with power law degree distribution follows a power law distribution. We find that the line graph of a hypergraph with power law degree distribution also has a power law degree distribution. As stated before, the degree of a node in a hypergraph is actually the size of the corresponding community. Therefore, our idea to construct a network with both power law individual (node degree) distribution and power law community size distribution is to generate the line graph of a hypergraph with power law degree distribution. To generate a hypergraph with power law degree distribution, we introduce preferential attachment to grow our hypergraphs.

To avoid confusion, we call the nodes of our hypergraphs directly the communities. We add *step by step* new communities and new hyperlinks to the starting small size hypergraph, namely a seed, which has no impact on our hypergraph model. Our hypergraph model is described by the following procedure:

1. Start with a seed hypergraph $H_0(N_0, L_0)$ with N_0 communities and L_0 hyperlinks.
2. Suppose that the desired number of individuals (hyperlinks) of the network to be generated is $L + L_0$. Determine the membership numbers for the L new hyperlinks: $M = \begin{bmatrix} m_1 & m_2 & \cdots & m_L \end{bmatrix}$. Note that the membership number vector M is the input parameter of our hypergraph model.
3. At growing step j, $j = 1, 2, \cdots, L$, add a new hyperlink j and a new community to the hypergraph. Make the new hyperlink j and the new community incident, and the membership number of j becomes 1.
   (a) Connect the new hyperlink j to the existing community i with probability $p_i = s_i / \sum s_i$, where s_i is the community size of i and $\sum s_i$ is the sum of community sizes of all the existing communities.
   (b) Repeat 3a) $m_j - 1$ times so that the membership number of the hyperlink j increases to m_j.
4. Repeat 3) until the number of hyperlinks increases to $L + L_0$.

Compute the metrics d_j, m_j, $\alpha_{i,j}$, s_j,u_j and $\beta_{i,j}$ using the methods given in Section 3.2 including the formulas (1) to (5). The membership number m_j of j at growing step j is a free parameter that we can tune.

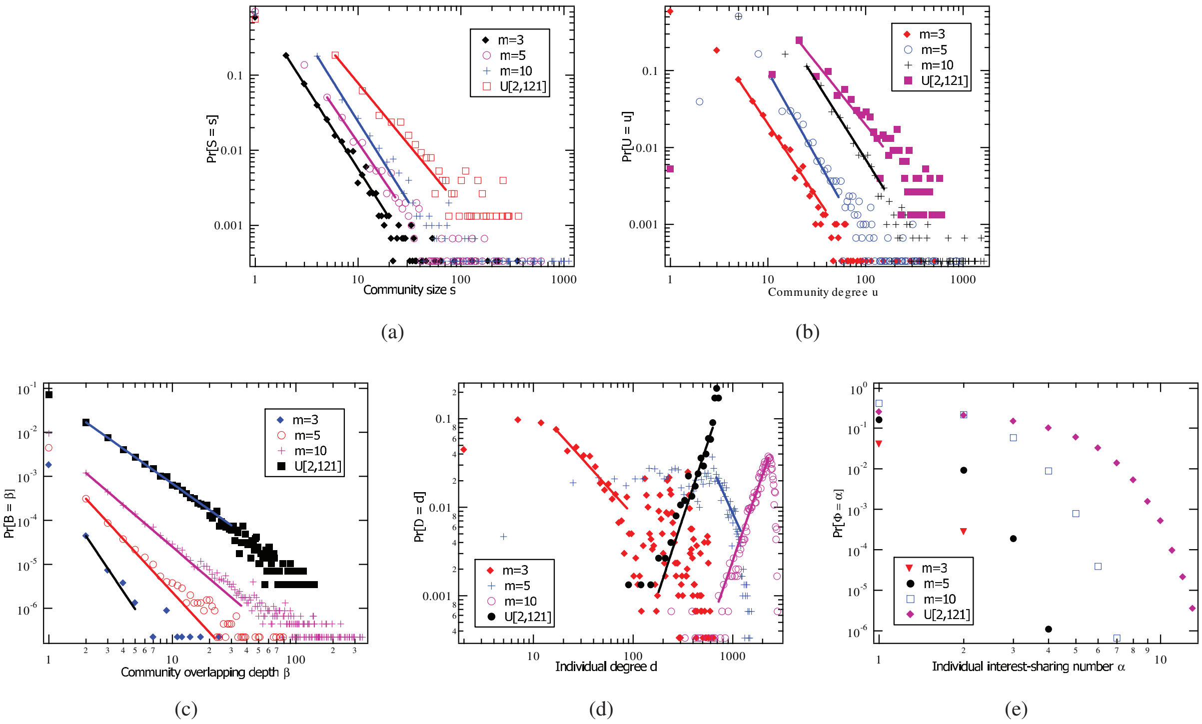


Figure 3: The probability density distribution of (a) community size s, (b) community degree u, (c) overlapping depth β, (d) individual degree d and (e) interest-sharing number α for $H_3\,(N, L)$, $H_5\,(N, L)$, $H_{10}\,(N, L)$ and $H_{U[2,121]}\,(N, L)$, where $N = 3015$ and $L = 3010$.

5. Simulation Results

5.1. Our hypergraph model

We use a linear hypergraph $H\,(15, 10)$ with the membership number $m_j = 3,\ j = 1, 2, \cdots 10$, as the starting seed. In our hypergraph model, we add 3000 new hyperlinks (individuals) and 3000 new nodes (communities) to the starting hypergraph by 3000 growing steps. Hence, all the hypergraphs we generate have 3015 nodes and 3010 hyperlinks.

During the 3000 growing steps, we apply the constant membership number $m_j = 3,\ j = 1, 2, \cdots, 3000$, and obtain the hypergraph H_3. Similarly, we generate H_5 and H_{10}. Since Palla et al. [12] have reported a power law membership number distribution in real-world networks, we apply power law distributed membership numbers and obtain the hypergraph H_{pow}. We generate a scale-free simple graph with 3000 nodes of which the degree D of a random node obeys $\Pr[D = d] = 2.67d^{-2.31}$, and take the nodal degree sequence as the membership numbers in the growing steps for the generation of H_{pow}. The maximum membership number of H_{pow} is 126. We also generate the hypergraph $H_{U[2,121]}$ with a uniformly distributed membership number in the interval $[2, 121]$, whose maximum membership number is close to that of H_{pow}.

We denote the community size and community degree of a random community by S and U, the community overlapping depth of a random pair of communities by B, the individual degree of a random individual by D, and the interest-sharing number of a random pair of hyperlinks by Φ.

Due to the preferential attachment, we expect the community sizes of H_3, H_5, H_{10}, $H_{U[2,121]}$ and H_{pow} to follow power law distributions, which are confirmed by Figure 3a and 4a. In Figure 3b and 4a the probability density functions (pdfs) of the community degrees of H_3, H_5, H_{10}, $H_{U[2,121]}$ and H_{pow} are all well fitted by power law functions with exponents shown in Table 2. The community overlapping depths of all hypergraphs follow power law distributions with exponents which are relatively larger in absolute values. The community overlapping depths of all hypergraphs are generally much smaller than the community size. The pdfs of the individual interest-sharing numbers of H_3, H_5, H_{10} and $H_{U[2,121]}$ can not be fitted by power law functions, while α of H_{pow} is well fitted by a power law function, as

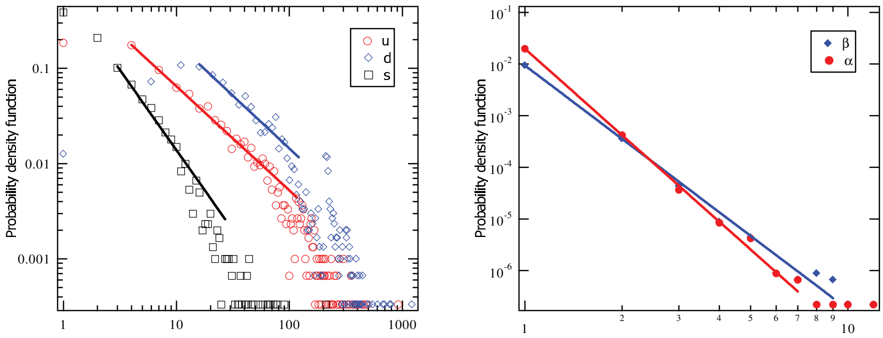


Figure 4: The probability density distribution of (a) community degree u, individual degree d and community size s, (b) individual interest-sharing number α and community overlapping depth β for $H_{pow}(3015, 3010)$ of which the membership number following a power-law with exponent of -2.31.

shown in Figure 3e and 4b. The pdfs of individual degrees of H_3, H_5 and H_{pow} are fitted by power law functions with negative exponents while those of H_{10} and $H_{U[2,121]}$ have positive exponents.

5.2. Real-world networks

5.2.1. The Hyves social network

The popular online social networking site in the Netherlands, Hyves, has more than 10 million users. Nearly half of the Hyves users make their profiles open to the public. From the open profiles we can see information of users including companies, schools, colleges, clubs and other organizations, to which they belong. According to the data we have collected, there are 17,619 users claiming that they belong to one or multiple schools. The total number of these claimed schools are about 10,326. We construct a network with users as nodes and two users are connected by a link when they belong to the same school, which implies that the members of a community are fully connected to each other in this extracted Hyves network.

In Figure 5a and 5b, we observe that the individual membership number m and community overlapping depth β and individual interest-sharing number α of the Hyves network generally smaller than the other three metrics, and all follow power law distributions with exponents of -4.83, -5.54 and -4.60 respectively. The pdfs of the community size s and the community degree u are well fitted by power functions with exponents of -1.87 and -2.43 respectively, while the pdfs of the individual degree d can only be fitted by a power law function with exponents of -1.29 in the middle region. Note that the exponents of m, α and β are more negative than those of s, u and d. The Hyves network has a rather high clustering coefficient $C = 0.84$, a high assortativity $\rho_D = 0.29$ and an short average path length $l = 6.73$.

5.2.2. The SourceForge software developers' network

SourceForge is a web-based project repository assisting developers to develop and distribute open source projects. SourceForge facilitates developers by providing a centralized storage and tools to manage the projects. Each project has multiple developers. We create a network in the same way that the Hyves network is constructed: take developers as nodes and connect two nodes if they belong to the same project: This implies that all the communities are cliques. Compared to the network obtained from Hyves, the network obtained from SourceForge has a much larger size of 161,653 nodes and 78,676 links.

As shown in Figure 5c, the pdfs of all the six metrics d_j, m_j, $\alpha_{i,j}$, s_j,u_j and $\beta_{i,j}$ of are well fitted by power law functions with exponents shown in Table 2. The SourceForge network also has a high clustering coefficient, a high assortativity coefficient and an small average path length, which are shown in Table 2.

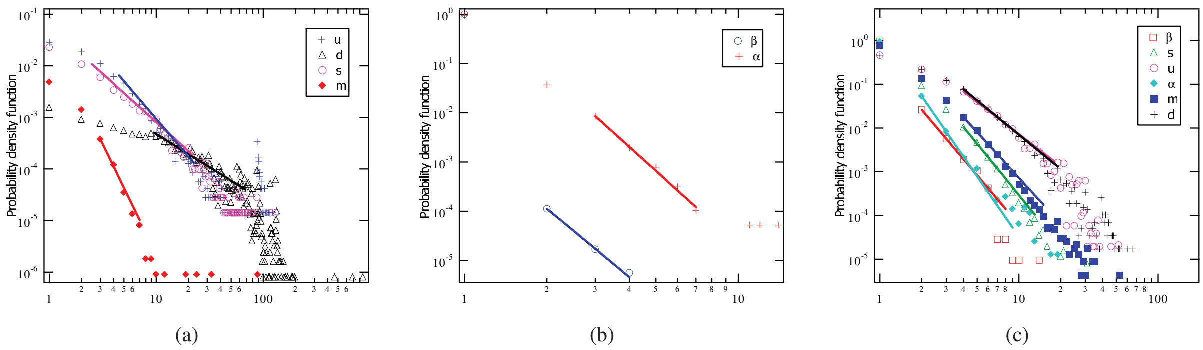


Figure 5: The probability density distribution of (a) community degree u, individual degree d, community size s and membership number m, (b) individual interest-sharing number α and community overlapping depth β for Hyves network. (c) The probability density distribution of u, d, s, m, α and β for SourceForge network.

5.2.3. Results of Palla et al. [12]

They defined *k-clique-community* as a union of all *k-clique* that can be reached from each other through a series of adjacent k-clique with adjacency means sharing $k-1$ nodes. Based on the concept of a *k-clique-community*, they studied the overlapping community structure of three large networks: the coauthorship network of the Los Alamos cond-mat archive ($k = 6$), the word association network of the South Florida Free Association norms ($k = 4$) and the protein interaction network of the yeast *S. Cerevisiae* from the DIP database ($k = 4$). They reported that the community size s of these three networks follow a power law distribution. The community overlapping depths and the membership numbers are in general small, and follow power law distributions with more negative exponents.

5.3. Discussion

The membership number m of the Hyves network, the SourceForge network, the coauthorship network, the word association network and the protein interaction network all follow a power law distribution and this is the motivation why we tune the membership number distribution of our hypergraph model to a power law distribution. Imposing a power law distribution to m in our hypergraph model makes the community degree, community size and the individual degree of our hypergraph model all follow a power law distribution as shown in Figure 4a. Consequently, the community overlapping depth β and the individual interest-sharing number α of our hypergraph model also follow power laws, as depicted and fitted in Figure 4b. We observe in Figure 4a and 4b that in general β and α are much smaller than m, s, u and d, and this is consistent with the observations of the mentioned five real-world networks in Figure 5a, 5b, 5c and figures of [12]. The tail of pdf of individual degree d of H_{pow} decreases more rapidly which is consistent with observations in Figure 5a and 5c.

Tuning the membership number m of our hypergraph model to constant numbers or a uniform distribution, does not influence the pdf of the group size s, due to the preferential attachment, as confirmed by Figure 3a. The pdfs of community degree u and community overlapping depth β still follow a power law distribution in H_3, H_5, H_{10} and $H_{U[2,121]}$. We conjecture that this phenomenon is caused by the power law distribution of the community size. In Figure 3d, we fit the middle region of the pdf of the individual degree d by a power function, although a large part of the pdf of d does not seem to be power-law-like. Interestingly, the exponent becomes positive in H_{10} and $H_{U[2,121]}$. The individual interest-sharing number α follows a non-power-law distribution shown in Figure 3e. Since $\alpha \leq m$ by definition, we only see a few points in the pdf of α for constant $m = 3, 5, 10$.

Table 2 reports the exponents of the power law distribution of these six metrics and also the clustering coefficient C, assortativity ρ_D and average shortest path length l. Our hypergraph model possesses high clustering coefficient, positive assortativity, short average path length, and these properties are independent of the input membership numbers m. Note that the exponents of the power law distributions of s, α and β of Hyves network and H_{pow} are quite close.

Table 2: Comparison of the meassures of our hypergraph model and the Hyves network and the SourceForge network. The measured metrics are: exponents of power law distribution of d, m, α, s, u and β, clustering coefficient C, assortativity coefficient ρ_D [8] and average path length l.

Network	Exponents of power law distribution						C	ρ_D	l
	m	d	u	s	α	β			
H_3	N.A.	−1.24	−1.90	−2.16	N.A.	−4.18	0.61	0.39	4.76
H_5	N.A.	−2.47	−2.25	−2.01	N.A.	−3.04	0.49	0.41	3.05
H_{10}	N.A.	3.25	−2.01	−2.19	N.A.	−2.41	0.76	0.62	4.02
$H_{U[2,121]}$	N.A.	3.41	−1.59	−1.67	N.A.	−1.99	0.44	0.58	5.22
H_{pow}	−2.31	−1.11	−1.09	−1.68	−5.56	−4.60	0.57	0.72	4.82
Hyves	−4.83	−1.29	−2.43	−1.87	−5.54	−4.60	0.84	0.29	6.73
SourceForge	−3.48	−2.61	−2.45	−3.91	−4.60	−3.76	0.63	0.40	7.06

6. Conclusion

Many real-world networks, especially social networks, exhibit an overlapping community structure. We present formulas which facilitate the computation for characterizing the overlapping community structure of networks. We have established a hypergraph-based social network model which exhibits innate tunable overlapping community structure. Our hypergraph representation of networks features the following properties: (i) facilitation of the computations of the characterizing metrics d_j, m_j, $\alpha_{i,j}$, s_j,u_j and $\beta_{i,j}$; (ii) facilitation in proving Theorem 3; (iii) ease in manipulating/tuning the overlapping community structure. By comparing simulation results of our model with results of the Hyves network, the SourceForge Network and results of Palla et al. [12], we have shown that our hypergraph model exhibits the common properties of large social networks: the community size s, the community degree u and the community overlapping depth β all follow a power law distribution, and our model possesses high clustering coefficient, positive assortativity, short average path length. By tuning the input individual membership number to follow a power law distribution, the individual degree d and the interest-sharing number α also follow a power law distribution.

7. Acknowledgement

This research was supported by the Next Generation Infrastructures foundation (http://www.nginfra.nl)

References

[1] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, Reviews of modern physics 74 (2002) 47–96.
[2] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Complex networks: Structure and dynamics, Physics Reports 424 (2006) 175–308.
[3] M. Girvan, M. E. J. Newman, Community structure in social and biological networks, Proceedings of the National Academy of Sciences of the United States of America 99 (12) (2002) 7821–7826.
[4] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, Science (New York, N.Y.) 286 (5439) (1999) 509–512.
[5] D. J. Watts, S. H. Strogatz, Collective dynamics of 'small -world' networks, Nature 393 (1998) 440–442.
[6] Y.-Y. Ahn, J. P. Bagrow, S. Lehmann, Link communities reveal multiscale complexity in networks, Nature 466 (7307) (2010) 761–764.
[7] M. E. J. Newman, Mixing patterns in networks, Phys. Rev. E 67 (2) (2003) 026126.
[8] P. Van Mieghem, H. Wang, X. Ge, S. Tang, F. A. Kuipers, Influence of assortativity and degree-preserving rewiring on the spectra of networks, The European Physical Journal B - Condensed Matter and Complex Systems (2010) 643–652.
[9] S. Fortunato, Community detection in graphs, Physics Reports 486 (3-5) (2010) 75 – 174.
[10] M. E. J. Newman, M. Girvan, Finding and evaluating community structure in networks, Phys. Rev. E 69 (2) (2004) 026113. doi:10.1103/PhysRevE.69.026113.
[11] P. Van Mieghem, X. Ge, P. Schumm, S. Trajanovski, H. Wang, Spectral graph analysis of modularity and assortativity, Phys. Rev. E 82 (5) (2010) 056113.
[12] G. Palla, I. Derenyi, I. Farkas, T. Vicsek, Uncovering the overlapping community structure of complex networks in nature and society, Nature 435 (7043) (2005) 814–818.
[13] T. Evans, R. Lambiotte, Overlapping communities, link partitions and line graphs, Proceedings of the European Conference on Complex Systems '09.

[14] A. McDaid, N. J. Hurley, Detecting highly overlapping communities with Model-based Overlapping Seed Expansion, in: ASONAM 2010, 2010.
[15] P. Pollner, G. Palla, T. Vicsek, Preferential attachment of communities: the same principle, but a higher level, EUROPHYS.LETT. 73 (2006) 478.
URL doi:10.1209/epl/i2005-10414-6
[16] M. E. J. Newman, D. J. Watts, S. H. Strogatz, Random graph models of social networks, Proc. Natl. Acad. Sci. USA 99 (2002) 2566–2572.
[17] B. Skyrms, R. Pemantle, A dynamic model of social network formation, Proceedings of the National Academy of Sciences of the United States of America 97 (16) (2000) 9340–9346.
[18] R. Toivonen, J.-P. Onnela, J. Saramki, J. Hyvnen, K. Kaski, A model for social networks, Physica A: Statistical and Theoretical Physics 371 (2) (2006) 851 – 860.
[19] J. Nacher, T. Yamada, S. Goto, M. Kanehisa, T. Akutsu, Two complementary representations of a scale-free network, Physica A: Statistical Mechanics and its Applications 349 (1-2) (2005) 349 – 363.
[20] A. Manka-Krason, A. Mwijage, K. Kulakowski, Clustering in random line graphs, Computer Physics Communications 181 (1) (2010) 118–121.
[21] P. Van Mieghem, Graph Spectra for Complex Networks, Cambridge University Press (Cambridge, U.K.), 2011.
[22] D. Cvetković, P. Rowlinson, S. K. Simić, Eigenvalue bounds for the signless laplacians, Publ. Inst. Math. (Beograd) 81 (95) (2007) 11–27.

AppendixA. Spectral property of networks described by hypergraph model

Theorem 3. *Denoting by λ an arbitrary eigenvalue of the adjacency matrix of the line graph of the hypergraph representing a network, we have*

$$\lambda \geq -m_{\max} \tag{A.1}$$

where $m_{\max}$ is the maximum membership number of that network.

Proof 4. *(a) Networks described by m-uniform hypergraph.*

With constant membership number m, these networks are represented by m-uniform hypergraphs $H_m(N, L)$, whose unsigned incidence matrix R has exactly m one-entries *and $N - m$* zero-entries *in each column. Thus, all the diagonal entries of $R^T R$ are m. The adjacency matrix of the line graph of $H_m(N, L)$ can be written as,*

$$A^{l(H_m)} = R^T R - mI \tag{A.2}$$

where $R^T R$ is a Gram matrix [21][22]. For all matrices $A_{n\times m}$ and $B_{m\times n}$ with $n \geq m$, it holds that $\lambda(AB) = \lambda(BA)$ and $\lambda(AB)$ has $n - m$ extra zero eigenvalues

$$\lambda^{n-m} \det(BA - \lambda I) = \det(AB - \lambda I)$$

which yields,

$$\det\left(A^{l(H_m)} - (\lambda - m) I\right) = \lambda^{L-N} \det\left(\left(RR^T\right)_{N\times N} - \lambda I\right)$$

The adjacency matrix $A^{l(H_m)}$ has at least $L - N$ eigenvalues $-m$. Since $x^T\left(R^T R\right)x = (Rx)^T Rx = \|Rx\|_2^2 \geq 0$ and $x^T\left(RR^T\right)x = \left(R^T x\right)^T R^T x = \left\|R^T x\right\|_2^2 \geq 0$, both $\left(R^T R\right)_{L\times L}$ and $\left(RR^T\right)_{N\times N}$ are positive semidefinite, hence all eigenvalues of $\left(R^T R\right)_{L\times L}$ are non-negative. Due to (A.2), the adjacency eigenvalues of $A^{l(H_m)}$ are not smaller than $-m$.

(b) Networks described non-uniform hypergraph.

We represent these networks by non-uniform hypergraphs $H(N, L)$ with maximum membership number $m_{\max}$. The matrix R of $H(N, L)$ has at most $m_{\max}$ one-entries *in each column. Therefore, the largest diagonal entry of $R^T R$ is $m_{\max}$. The adjacency matrix of the line graph of non-uniform hypergraph $H(N, L)$ is,*

$$A_{l(H)} = R^T R + C - m_{\max} I \tag{A.3}$$

where $C = diag\begin{pmatrix} c_{11} & c_{22} & \cdots & c_{LL} \end{pmatrix}$ and $c_{jj} \geq 0$, $1 \leq j \leq L$. By adding C to $R^T R$, we make all the diagonal entries of $R^T R + C$ equal to $\beta_j^{\max}$. Since $x^T\left(R^T R + C\right)x = x^T\left(R^T R\right)x + x^T\left(\sqrt{C}^T \sqrt{C}\right)x = \|Rx\|_2^2 + \left\|\sqrt{C}x\right\|_2^2 \geq 0$, $R^T R + C$ is also positive semidefinite, where $x_{L\times 1}$ is an arbitrary vector and $\sqrt{C} = diag\begin{pmatrix} \sqrt{c_{11}} & \sqrt{c_{22}} & \cdots & \sqrt{c_{LL}} \end{pmatrix}$. Hence, $-m_{\max} \leq \lambda_{\min}$.