A High-Order Proximity-Incorporated Nonnegative Matrix Factorization-based Community Detector Supplementary File

Zhigang Liu, Yugen Yi, *Member, IEEE*, Xin Luo, *Senior Member*, *IEEE*

I. Introduction

This is the supplementary file for the paper entitled "*A High-Order Proximity-Incorporated Nonnegative Matrix Factorization-based Community Detector*". We have put tables of main acronyms and symbols in Section II, the proof of *Theorem* 1 and *Convergence Proof* in Section III, and some figures and tables of experimental results in Section IV.

II. Acronyms and Symbols

Main acronyms and symbols of the paper are summarized in Tables S.I and S.II.

III. Proof of Theorems 1 and Convergence Proof

A. Proof of Theorem 1

Proof. We prove the correctness of the learning rules in (23) in the manuscript separately in the following.

a) Updating *X*: If the learning rule in (23a) in the manuscript converges, then we achieve $X^{(\infty)}=X^{(t+1)}=X^{(t)}=X^{(*)}$, where t denotes the t-th iteration. Thus, for each x_{ik} in X, we have the following inference:

$$
\begin{aligned}
x_{ik}^{(*)} &= \lim_{t\to\infty} x_{ik}^{(t)} = x_{ik}^{(t+1)} \\
&= x_{ik}^{(t)}\left(1-\beta+\beta\frac{\left(\left(1+\lambda\right)\tilde{A}X^{(t)}+\vartheta Y+\vartheta U\right)_{sk}}{\left(X^{(t)}\left(X^{(t)}\right)^{\mathrm{T}}X^{(t)}+2\vartheta X^{(t)}+\lambda\tilde{D}X^{(t)}\right)_{ik}}\right) \\
&\Rightarrow x_{ik}^{(*)}=0,\ \text{or}\ \left(X^{(*)}\left(X^{(*)}\right)^{\mathrm{T}}X^{(*)}+2\vartheta X^{(*)}+\lambda\tilde{D}X^{(*)}-\left(1+\lambda\right)\tilde{A}X^{(*)}-\vartheta Y-\vartheta U\right)_{ik}=0,
\end{aligned}
\tag{S1}
$$

which results in

$$
\left(X^{(*)}\left(X^{(*)}\right)^{\mathrm{T}}X^{(*)}+2\vartheta X^{(*)}+\lambda\tilde{D}X^{(*)}-\left(1+\lambda\right)\tilde{A}X^{(*)}-\vartheta Y-\vartheta U\right)_{ik}x_{ik}^{(*)}=0,
\tag{S2}
$$

which is identical with (22a) in the manuscript.

b) Updating *Y*: If the learning rule in (23b) converges, then we achieve $Y^{(\infty)}=Y^{(t+1)}=Y^{(t)}=Y^{(*)}$, where t denotes the t-th iteration. Thus, for each y_{jk} in Y, we have the following inference:

$$
\begin{aligned}
y_{jk}^{(*)} &= \lim_{t\to\infty} y_{jk}^{(t)} = y_{jk}^{(t+1)} = y_{jk}^{(t)}\frac{\left(\tilde{A}U+X\right)_{jk}}{\left(Y^{(t)}U^{\mathrm{T}}U+Y^{(t)}\right)_{jk}} \\
&\Rightarrow y_{jk}^{(*)}=0,\ \text{or}\ \left(Y^{(*)}U^{\mathrm{T}}U+Y^{(*)}-\tilde{A}U-X\right)_{jk}=0,
\end{aligned}
\tag{S3}
$$

TABLE S.I
MAIN ACRONYM APPOINTMENT

Acronym	Description
IoT	Internet of Things
SNMF	Symmetric and Nonnegative Matrix Factorization
LF	Latent Factor
FOP	First-Order Proximity
PMI	Pointwise Mutual Information
HOP	High-Order Proximity
HOP-NMF	High-Order Proximity-incorporated, Nonnegative Matrix Factorization
INE	Iterative Network Enhancement
CGF	Capacity-enlarged and Graph-regularized Factorization
KKT	Karush-Kuhn-Tucker
NMU	Nonnegative Multiplicative Update
DFS	Depth-First Search
NMI	Normalized Mutual Information
AC	Clustering Accuracy

TABLE S.II
SYMBOL APPOINTMENT

Symbol	Description
G	Target graph describing an undirected and unweighted network.
V, E	Node set and edge set of G, respectively.
v_i, e_{ij}	Single node and edge in V and E, respectively.
n, m	Number of nodes in V and number of edges in E, respectively.
C_l	Community l.
K	Number of communities.
$A^{n\times n}$, $\hat{A}^{n\times n}$	Adjacency matrix and its low-rank approximation.
$\tilde{A}$	Adjacency matrix of an HOP-enhanced network.
a_{ij}, $\hat{a}_{ij}$, $\tilde{a}_{ij}$	Single elements in A, $\hat{A}$ and $\tilde{A}$.
W, $\tilde{W}$	Weight matrix related to A and $\tilde{A}$.
D, $\tilde{D}$	Diagonal matrix related to A and $\tilde{A}$.
L, $\tilde{L}$	Laplacian matrix related to A and $\tilde{A}$.
$X^{n\times K}$, $Y^{n\times K}$	Latent factor matrices of A.
$U^{n\times K}$	Bridging matrix related to X and Y.
x_{ik}, y_{jk}, u_{sk}	Single entries in X, Y and U.
$x_{ik}^{(t)}$, $y_{jk}^{(t)}$, $u_{sk}^{(t)}$	State values of x_{ik}, y_{jk} and u_{sk} in the t-th iteration.
Φ, Ψ, K	Lagrangian multiplier matrices related to X, Y and U.
ϕ_{ik}, ψ_{jk}, κ_{sk}	Single entries in Φ, Ψ and K.
ξ, ζ, ϑ	Equality-regularization coefficients
λ	Graph regularization coefficient.
β	Adjusting coefficient.
r, d	Order and count of HOP-based INE.
H_k	A set of k-th-order node pairs.
ω_k	Weight for the k-th-order connection of nodes
ε	Threshold coefficient.
P	HOP matrix.
$P(v_i, v_j)$	HOP index between nodes v_i and v_j.
$p(v_i)$	The probability of observing v_i.
$p(v_i, v_j)$	The probability of observing v_i and v_j together.
$N_k(v_i)$, $N_k(v_j)$	Times that node v_i (v_j) appears in a k-th-order node pair set.
$N_k(v_i, v_j)$	Times that node pair (v_i, v_j) appears in a k-th-order node pair set.
$\cup$, $\cap$	Union and intersection set operation, respectively.
$\varnothing$	Null set.
$\lvert\cdot\rvert$	Cardinality of an enclosed set.
$\lVert\cdot\rVert_F$	Frobenius norm of an enclosed matrix
$\mathrm{Tr}(\cdot)$	Trace of an enclosed matrix.

which results in

$$\left(Y^{(*)}U^{\mathrm{T}}U+Y^{(*)}-\tilde{A}U-X\right)_{jk}y_{jk}^{(*)}=0. \tag{S4}$$

Note that (S4) is identical with (22b) in the manuscript.

c) Updating *U*: If the learning rule in (23c) in the manuscript converges, then we achieve $U^{(\infty)}=U^{(t+1)}=U^{(t)}=U^{(*)}$, where t denotes the t-th iteration. Thus, for each u_{sk} in U, we have the following inference:

$$\begin{aligned} &u_{sk}^{(*)}=\lim_{t\to\infty}u_{sk}^{(t)}=u_{sk}^{(t+1)}=u_{sk}^{(t)}\frac{\left(\tilde{A}^{\mathrm{T}}Y+X\right)_{sk}}{\left(U^{(t)}Y^{\mathrm{T}}Y+U^{(t)}\right)_{sk}} \\ &\Rightarrow u_{sk}^{(*)}=0,\ \text{or}\ \left(U^{(*)}Y^{\mathrm{T}}Y+U^{(*)}-\tilde{A}^{\mathrm{T}}Y-X\right)_{sk}=0, \end{aligned} \tag{S5}$$

which results in

$$\left(U^{(*)}Y^{\mathrm{T}}Y+U^{(*)}-\tilde{A}^{\mathrm{T}}Y-X\right)_{sk}u_{sk}^{(*)}=0, \tag{S6}$$

which is identical with (22c) in the manuscript.

Based on the above analysis, we conclude that converging solutions of X, Y and U satisfy the KKT optimality conditions.■

B. Convergence Proof

Following [36, 47, 48], we conduct the proof by introducing auxiliary functions for the objective function of optimization problem (19) in the manuscript to help us prove that J_{CGF} is non-increasing during the updating of each single variable. We start the proof by recalling the definition of an auxiliary function [36, 47].

***Definition* S1.** Given a function $\Gamma(x, x^{(t)})$, if it simultaneously fulfills that $\Gamma(x, x^{(t)})\geq F(x)$ and $\Gamma(x, x)\geq F(x)$, then $\Gamma(x, x^{(t)})$ can be an auxiliary function for $F(x)$.

Based on *Definition* S1, we recall the property of an auxiliary function [36, 47].

***Lemma* S1.** If $\Gamma(x, x^{(t)})$ is an auxiliary function of $F(x)$, then $F(x)$ is non-increasing in each iteration by the following learning rule:

$$x^{(t+1)}=\arg\min_{x}\Gamma\left(x,x^{(t)}\right). \tag{S7}$$

Note that the proof for *Lemma* S1 is provided in [47]. Thus, we aim to prove that (23) in the manuscript for X, Y and U is essentially equivalent to (S7) with properly-designed auxiliary functions.

Due to the non-convexity of J_{CGF}, we aim to prove its non-increment with separately X, Y or U under (23) in the manuscript by keeping the other two alternatively fixed, which is consistent with the proof sketch in [36, 47].

a) Updating *X*: Considering an arbitrary entry x_{ik} in X, we adopt $F(x_{ik})$ to denote the corresponding component from J_{CGF} that is relevant to x_{ik} only. We calculate the first-order and second-order derivatives of $F(x_{ik})$ with respect to x_{ik}, i.e.,

$$F'\left(x_{ik}\right)=\frac{\partial J_{CGF}}{\partial x_{ik}}=\left(-\vartheta Y-\vartheta U+XX^{\mathrm{T}}X-\tilde{A}X+\lambda\tilde{L}X\right)_{ik}+2\vartheta x_{ik}, \tag{S8}$$

$$F''\left(x_{ik}\right)=\frac{\partial^2 J_{CGF}}{\partial x_{ik}^2}=\left(X^{\mathrm{T}}X\right)_{kk}+\left(XX^{\mathrm{T}}-\tilde{A}+\lambda\tilde{L}\right)_{ii}+2\vartheta+x_{ik}^2. \tag{S9}$$

***Lemma* S2.** The following $\Gamma(x_{ik}, x_{ik}^{(t)})$ is an auxiliary function for $F(x_{ik})$.

$$\Gamma\left(x_{ik},x_{ik}^{(t)}\right)=F\left(x_{ik}^{(t)}\right)+F'\left(x_{ik}^{(t)}\right)\left(x_{ik}-x_{ik}^{(t)}\right)+\frac{\left(XX^{\mathrm{T}}X+\lambda\tilde{D}X\right)_{ik}+2\vartheta x_{ik}^{(t)}}{2\beta x_{ik}^{(t)}}\left(x_{ik}-x_{ik}^{(t)}\right)^2. \tag{S10}$$

Proof. Note that with (S10) $\Gamma(x_{ik}, x_{ik})=F(x_{ik})$ evidently holds. Thus, we aim to prove that $\Gamma(x_{ik}, x_{ik}^{(t)})\geq F(x_{ik})$. To achieve this objective, we expand $F(x_{ik})$ to its second-order Taylor series at the state point $x_{ik}^{(t)}$ as:

$$F\left(x_{ik}\right)=F\left(x_{ik}^{(t)}\right)+F'\left(x_{ik}^{(t)}\right)\left(x_{ik}-x_{ik}^{(t)}\right)+\frac{1}{2}\left(\left(X^{\mathrm{T}}X\right)_{kk}+\left(XX^{\mathrm{T}}-\tilde{A}+\lambda\tilde{L}\right)_{ii}+2\vartheta+x_{ik}^{2}\right)\left(x_{ik}-x_{ik}^{(t)}\right)^{2}. \quad \text{(S11)}$$

By combining (S10) and (S11), we see that the desired condition of $\Gamma(x_{ik}, x_{ik}^{(t)})\geq F(x_{ik})$ is equivalent to the following condition:

$$\frac{\left(XX^{\mathrm{T}}X+\lambda\tilde{D}X\right)_{ik}+2\vartheta x_{ik}^{(t)}}{\beta x_{ik}^{(t)}}\geq\left(X^{\mathrm{T}}X\right)_{kk}+\left(XX^{\mathrm{T}}+\lambda\tilde{D}-\left(1+\lambda\right)\tilde{A}\right)_{ii}+2\vartheta+\left(x_{ik}^{(t)}\right)^{2} \quad \text{(S12)}$$

Since $x_{ik}^{(t)}\geq 0$[1], (S12) can be reduced into the following form:

$$\frac{1}{\beta}\left(XX^{\mathrm{T}}X+\lambda\tilde{D}X\right)_{ik}\geq x_{ik}^{(t)}\left(X^{\mathrm{T}}X\right)_{kk}+\left(XX^{\mathrm{T}}+\lambda\tilde{D}-\left(1+\lambda\right)\tilde{A}\right)_{ii}x_{ik}^{(t)}+\left(x_{ik}^{(t)}\right)^{3}. \quad \text{(S13)}$$

It is easy to show that

$$\left(XX^{\mathrm{T}}X\right)_{ik}=\sum_{f=1}^{K}\left(x_{if}\left(X^{\mathrm{T}}X\right)_{fk}\right)=\sum_{f=1,f\neq k}^{K}\left(x_{if}\left(X^{\mathrm{T}}X\right)_{fk}\right)+x_{ik}\left(X^{\mathrm{T}}X\right)_{kk}\geq x_{ik}\left(X^{\mathrm{T}}X\right)_{kk}. \quad \text{(S14)}$$

Similarly, we have

$$\left(XX^{\mathrm{T}}X\right)_{ik}\geq\left(XX^{\mathrm{T}}\right)_{ii}x_{ik}, \quad \text{(S15)}$$

and

$$\left(XX^{\mathrm{T}}X\right)_{ik}\geq x_{ik}^{3}. \quad \text{(S16)}$$

By combining (S15)-(S16), we have

$$\left(XX^{\mathrm{T}}X\right)_{ik}\geq\frac{1}{3}\left(x_{ik}\left(X^{\mathrm{T}}X\right)_{kk}+\left(XX^{\mathrm{T}}\right)_{ii}x_{ik}+x_{ik}^{3}\right). \quad \text{(S17)}$$

On the other hand, we have

$$\lambda\left(\tilde{D}X\right)_{ik}=\lambda\sum_{h=1}^{n}\tilde{D}_{ih}x_{hk}=\lambda\sum_{h=1,h\neq i}^{n}\tilde{D}_{ih}x_{hk}+\lambda\tilde{D}_{ii}x_{ik}\geq\lambda\tilde{D}_{ii}x_{ik}\geq\lambda\left(\tilde{D}-\left(1+\lambda\right)\tilde{A}\right)_{ii}x_{ik}. \quad \text{(S18)}$$

Based on the inferences in (S17) and (S18), we conclude that (S13) holds when $0<\beta\leq 1/3$, resulting in the establishment of (S12).

Note that it is hard to rigorously show that (S13) still holds when $1/3<\beta\leq 1$. But in fact, (S13) most likely would hold when $1/3<\beta\leq 1$, since in the shrinkage process of (S14)-(S18) we cast away many positive terms that ensures that (S13) holds. Besides, the empirical convergence analysis later has also confirmed this inference. Therefore, we have $\Gamma(x_{ik}, x_{ik}^{(t)})\geq F(x_{ik})$, which makes $\Gamma(x_{ik}, x_{ik}^{(t)})$ be an auxiliary function of $F(x_{ik})$.□

Based on *Lemmas* S1 and S2, we then have the following theorem.

***Theorem* S1.** The value of J_{CGF} in (19) in the manuscript keeps non-increasing when updating X by the learning rule (23a).

Proof. It is equivalent to proving that the learning rule in (S7) is consistent with that in (23a). Thus, by replacing $\Gamma(x_{ik}, x_{ik}^{(t)})$ in (S7) with the auxiliary function in (S10), we have:

[1] Please note that if $x_{ik}^{(t)}$ goes to zero after the t-th iteration, then it will keep zero constantly with the learning scheme (22) in the follow-up iterations, which will not affect the convergence of the learning algorithm. Therefore, without loss of generality, we only take into account the case of $x_{ik}^{(t)}>0$ in our proof. Similarly, only the case of $y_{jk}^{(t)}>0$ or $u_{sk}^{(t)}>0$ is taken into consideration when we present the proof of **Updating *Y*** and **Updating *U*** later, respectively.

$$
\begin{aligned}
&x_{ik}^{(t+1)} = \arg\min_{x_{ik}} \Gamma\left(x_{ik}, x_{ik}^{(t)}\right) \\
&\Rightarrow F'\left(x_{ik}^{(t)}\right) + \frac{\left(XX^{\mathrm{T}}X + \lambda\tilde{D}X\right)_{ik} + 2\vartheta x_{ik}^{(t)}}{\beta x_{ik}^{(t)}}\left(x_{ik} - x_{ik}^{(t)}\right) = 0 \\
&\Rightarrow x_{ik} = x_{ik}^{(t)} - x_{ik}^{(t)} \frac{\beta F'\left(x_{ik}^{(t)}\right)}{\left(XX^{\mathrm{T}}X + \lambda\tilde{D}X\right)_{ik} + 2\vartheta x_{ik}^{(t)}} \\
&\Rightarrow x_{ik}^{(t+1)} = x_{ik}^{(t)}\left(1 - \beta + \beta\frac{\left((1+\lambda)\tilde{A}X + \vartheta Y + \vartheta U\right)_{ik}}{\left(XX^{\mathrm{T}}X + \lambda\tilde{D}X + 2\vartheta X\right)_{ik}}\right),
\end{aligned}
\tag{S19}
$$

which is identical with the learning rule in (23a). Note that x_{ik} is an arbitrary entry of X. Hence, $\forall i\in\{1, 2, \ldots, n\}$, $k\in\{1, 2, \ldots, K\}$, $F(x_{ik})$ is non-increasing with the learning scheme (23a) as Y and U are alternatively fixed. Based on the inferences above, we conclude that J_{CGF} is non-increasing when updating X by the learning rule (23a). Hence, *Theorem* S1 stands.■

b) Updating *Y*: By analogy, considering an arbitrary entry y_{jk} in Y, we adopt the function $F(y_{jk})$ to denote the corresponding component from J_{CGF} that is relevant to y_{jk} only. Then, we have the first-order and second-order derivatives of $F(y_{jk})$ with respect to y_{jk}, i.e.,

$$F'\left(y_{jk}\right) = \frac{\partial J_{CGF}}{\partial y_{jk}} = \left(\vartheta YU^{\mathrm{T}}U - \vartheta\tilde{A}U - \vartheta X\right)_{jk} + \vartheta y_{jk}, \tag{S20}$$

$$F''\left(y_{jk}\right) = \frac{\partial^2 J_{CGF}}{\partial y_{jk}^2} = \vartheta\left(U^{\mathrm{T}}U\right)_{kk} + \vartheta. \tag{S21}$$

***Lemma* S3.** The following $\Gamma(y_{jk}, y_{jk}^{(t)})$ is an auxiliary function for $F(y_{jk})$.

$$\Gamma\left(y_{jk}, y_{jk}^{(t)}\right) = F\left(y_{jk}^{(t)}\right) + F'\left(y_{jk}^{(t)}\right)\left(y_{jk} - y_{jk}^{(t)}\right) + \frac{\vartheta\left(YU^{\mathrm{T}}U\right)_{jk} + \vartheta y_{jk}^{(t)}}{2y_{jk}^{(t)}}\left(y_{jk} - y_{jk}^{(t)}\right)^2. \tag{S22}$$

Proof. With (S18), $\Gamma(y_{jk}, y_{jk})=F(y_{jk})$ holds. Next, we aim to prove that $\Gamma(y_{jk}, y_{jk}^{(t)})\geq F(y_{jk})$. To do this, we expand $F(y_{jk})$ to its second-order Taylor series at the state point $y_{jk}^{(t)}$ as:

$$F\left(y_{jk}\right) = F\left(y_{jk}^{(t)}\right) + F'\left(y_{jk}^{(t)}\right)\left(y_{jk} - y_{jk}^{(t)}\right) + \frac{1}{2}\left(\vartheta\left(U^{\mathrm{T}}U\right)_{kk} + \vartheta\right)\left(y_{jk} - y_{jk}^{(t)}\right)^2. \tag{S23}$$

Combining (S22) and (S23), we see that the desired condition of $\Gamma(y_{jk}, y_{jk}^{(t)})\geq F(y_{jk})$ is equivalent to the following condition:

$$\left(\vartheta\left(YU^{\mathrm{T}}U\right)_{jk} + \vartheta y_{jk}^{(t)}\right)\Big/ y_{jk}^{(t)} \geq \vartheta\left(U^{\mathrm{T}}U\right)_{kk} + \vartheta. \tag{S24}$$

Since $y_{jk}^{(t)}\geq 0$ and $\vartheta>0$, (S24) can be reduced into the following form:

$$\left(YU^{\mathrm{T}}U\right)_{jk} \geq y_{jk}^{(t)}\left(U^{\mathrm{T}}U\right)_{kk}. \tag{S25}$$

To prove (S25), we have the following inferences:

$$\left(YU^{\mathrm{T}}U\right)_{jk} = \sum_{f=1}^{K} y_{jf}^{(t)}\left(U^{\mathrm{T}}U\right)_{fk} = \sum_{f=1, f\neq k}^{K} y_{jf}^{(t)}\left(U^{\mathrm{T}}U\right)_{fk} + y_{jk}^{(t)}\left(U^{\mathrm{T}}U\right)_{kk} \geq y_{jk}^{(t)}\left(U^{\mathrm{T}}U\right)_{kk}. \tag{S26}$$

Hence, (S24) holds based on (S26), thus making $\Gamma(y_{jk}, y_{jk}^{(t)})$ be an auxiliary function of $F(y_{jk})$.□

Based on *Lemmas* S1 and S3, we then have the following theorem.

***Theorem* S2.** The value of J_{CGF} in (19) in the manuscript keeps non-increasing when updating Y by the learning rule (23b).

Proof. It is equivalent to proving that the learning rule in (S7) is consistent with that in (23b). By replacing $\Gamma(y_{jk}, y_{jk}^{(t)})$ in (S7) with the auxiliary function in (S23), we have:

$$y_{jk}^{(t+1)}=\arg\min_{y_{jk}}\Gamma\left(y_{jk},y_{jk}^{(t)}\right)$$
$$\Rightarrow F'\left(y_{jk}^{(t)}\right)+\frac{\vartheta\left(YU^{\mathrm{T}}U\right)_{jk}+\vartheta y_{jk}^{(t)}}{y_{jk}^{(t)}}\left(y_{jk}-y_{jk}^{(t)}\right)=0 \tag{S27}$$
$$\Rightarrow y_{jk}=y_{jk}^{(t)}-y_{jk}^{(t)}\frac{F'\left(y_{jk}^{(t)}\right)}{\vartheta\left(YU^{\mathrm{T}}U\right)_{jk}+\vartheta y_{jk}^{(t)}}\Rightarrow y_{jk}^{(t+1)}=y_{jk}^{(t)}\frac{\left(\tilde{A}U+X\right)_{jk}}{\left(YU^{\mathrm{T}}U+Y\right)_{jk}},$$

which is identical with the learning rule in (23b). Similarly, $\forall j\in\{1, 2, \ldots, n\}$, $k\in\{1, 2, \ldots, K\}$, $F(y_{jk})$ keeps non-increasing under the learning scheme (23b) as X and U are alternatively fixed. Thus, we conclude that J_{CGF} keeps non-increasing when updating Y by the learning rule (23b), and *Theorem* S2 stands.■

c) Updating U: Similarly, considering an arbitrary entry u_{sk} in U, we adopt $F(u_{sk})$ to denote the corresponding component from J_{CGF} that is relevant to u_{sk} only. Then, we have the first-order and second-order derivatives of $F(u_{sk})$ with respect to u_{sk}, i.e.,

$$F'\left(u_{sk}\right)=\frac{\partial J_{CGF}}{\partial u_{sk}}=\left(\vartheta UY^{\mathrm{T}}Y-\vartheta\tilde{A}^{\mathrm{T}}Y-\vartheta X\right)_{sk}+\vartheta u_{sk}, \tag{S28}$$
$$F''\left(u_{sk}\right)=\frac{\partial^2 J_{CGF}}{\partial u_{sk}^2}=\vartheta\left(Y^{\mathrm{T}}Y\right)_{kk}+\vartheta. \tag{S29}$$

Next, we build an auxiliary function for $F(u_{sk})$.

***Lemma* S4.** The following $\Gamma(u_{sk}, u_{sk}^{(t)})$ is an auxiliary function for $F(u_{sk})$.

$$\Gamma\left(u_{sk},u_{sk}^{(t)}\right)=F\left(u_{sk}^{(t)}\right)+F'\left(u_{sk}^{(t)}\right)\left(u_{sk}-u_{sk}^{(t)}\right)+\frac{\vartheta\left(UY^{\mathrm{T}}Y\right)_{sk}+\vartheta u_{sk}^{(t)}}{2u_{sk}^{(t)}}\left(u_{sk}-u_{sk}^{(t)}\right)^2. \tag{S30}$$

Proof. Based on (S30), $\Gamma(u_{sk}, u_{sk})=F(u_{sk})$ holds. Next, we need to prove that $\Gamma(u_{sk}, u_{sk}^{(t)})\geq F(u_{sk})$. To do this, we expand $F(u_{sk})$ to its second-order Taylor series at the state point $u_{sk}^{(t)}$, i.e.,

$$F\left(u_{sk}\right)=F\left(u_{sk}^{(t)}\right)+F'\left(u_{sk}^{(t)}\right)\left(u_{sk}-u_{sk}^{(t)}\right)+\frac{1}{2}\left(\vartheta\left(Y^{\mathrm{T}}Y\right)_{kk}+\vartheta\right)\left(u_{sk}-u_{sk}^{(t)}\right)^2. \tag{S31}$$

By combining (S30) and (S31), we see that the desired condition of $\Gamma(u_{sk}, u_{sk}^{(t)})\geq F(u_{sk})$ is equivalent to the following condition:

$$\left(\vartheta\left(UY^{\mathrm{T}}Y\right)_{sk}+\vartheta u_{sk}^{(t)}\right)\Big/u_{sk}^{(t)}\geq\vartheta\left(Y^{\mathrm{T}}Y\right)_{kk}+\vartheta. \tag{S32}$$

Since $u_{sk}^{(t)}\geq 0$ and $\vartheta>0$, (S32) can be reduced into the following form:

$$\left(UY^{\mathrm{T}}Y\right)_{sk}\geq u_{sk}^{(t)}\left(Y^{\mathrm{T}}Y\right)_{kk}. \tag{S33}$$

To prove (S33), we have the following inferences:

$$\left(UY^{\mathrm{T}}Y\right)_{sk}=\sum_{f=1}^{K}u_{sf}^{(t)}\left(Y^{\mathrm{T}}Y\right)_{fk}=\sum_{f=1,f\neq k}^{K}u_{sf}^{(t)}\left(Y^{\mathrm{T}}Y\right)_{fk}+u_{sk}^{(t)}\left(Y^{\mathrm{T}}Y\right)_{kk}\geq u_{sk}^{(t)}\left(Y^{\mathrm{T}}Y\right)_{kk}. \tag{S34}$$

Hence, (S34) holds and we have $\Gamma(u_{sk}, u_{sk}^{(t)})\geq F(u_{sk})$, which makes $\Gamma(u_{sk}, u_{sk}^{(t)})$ be an auxiliary function of $F(u_{sk})$.□

Based on *Lemmas* S1 and S4, we can prove the convergence of the learning rule (23c) in the manuscript.

***Theorem* S3.** The value of J_{CGF} in (19) in the manuscript keeps non-increasing when updating U by the learning rule (23c).

Proof. It is equivalent to proving that the learning rule in (S7) is consistent with that in (23c). By replacing $\Gamma(u_{sk}, u_{sk}^{(t)})$ in (S7) with the auxiliary function in (S30), we have:

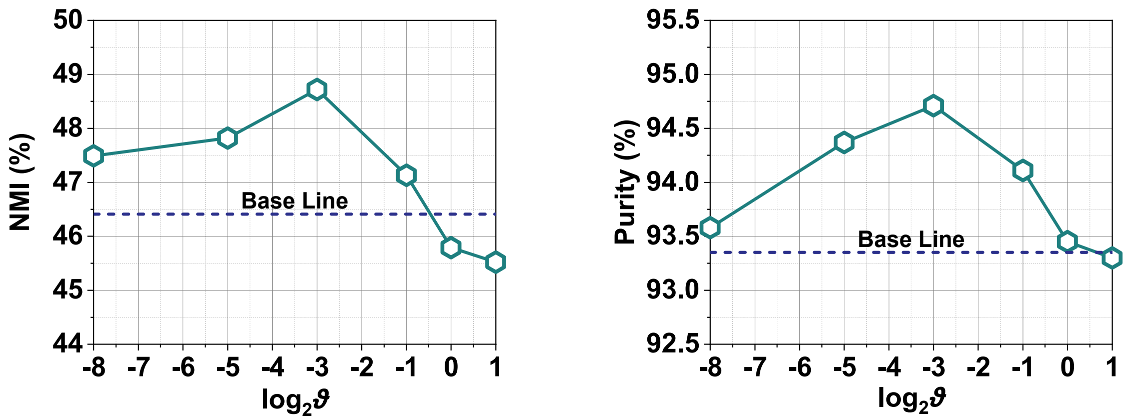
$$
\begin{aligned}
& u_{sk}^{(t+1)} = \arg\min_{u_{sk}} \Gamma\left(u_{sk}, u_{sk}^{(t)}\right), \\
& \Rightarrow F'\left(u_{sk}^{(t)}\right) + \frac{\vartheta\left(UY^{\mathrm{T}}Y\right)_{sk} + \vartheta u_{sk}^{(t)}}{u_{sk}^{(t)}}\left(u_{sk} - u_{sk}^{(t)}\right) = 0, \\
& \Rightarrow u_{sk} = u_{sk}^{(t)} - u_{sk}^{(t)} \frac{F'\left(u_{sk}^{(t)}\right)}{\vartheta\left(UY^{\mathrm{T}}Y\right)_{sk} + \vartheta u_{sk}^{(t)}}, \Rightarrow u_{sk}^{(t+1)} = u_{sk}^{(t)} \frac{\left(\tilde{A}^{\mathrm{T}}Y + X\right)_{sk}}{\left(UY^{\mathrm{T}}Y + U\right)_{sk}},
\end{aligned}
\tag{S35}
$$

which is identical with the learning rule in (23c). Similarly, $\forall s\in\{1, 2, \ldots, n\}$, $k\in\{1, 2, \ldots, K\}$, $F(u_{sk})$ is non-increasing with the learning scheme (23c) as X and Y are alternatively fixed. Thus, we conclude that J_{CGF} keeps non-increasing when updating U by the learning rule (23c), and *Theorem* S3 stands.■

With *Theorems* S1-S3, we conclude that the convergence of the HOP-NMF-based community detector with the learning scheme (23) in the manuscript is guaranteed. Besides, with *Theorem* 1 in the manuscript, we have that the solution sequences of the CGF algorithm fulfils the KKT conditions, thereby making it converge to KKT stationary points of its learning objective.

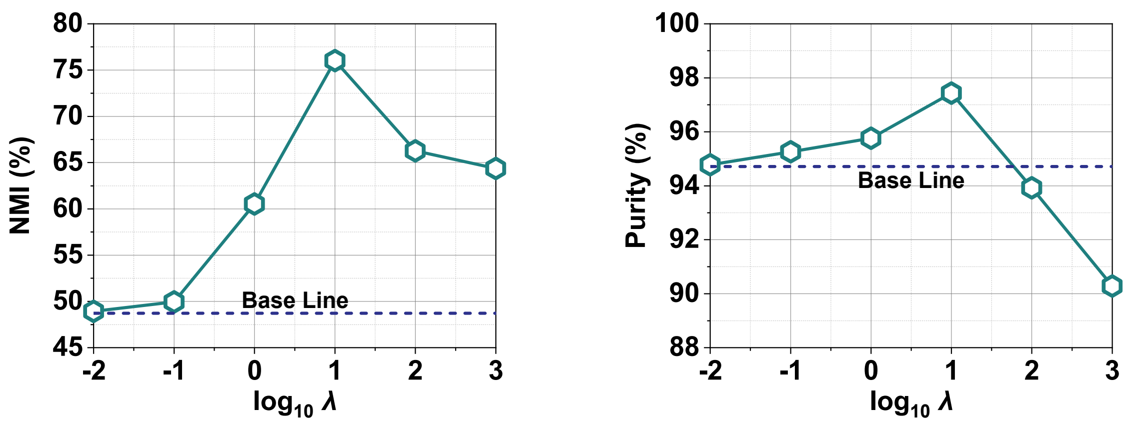
IV. Some Figures and Tables of Experimental Results

Some experimental results are shown in this section.



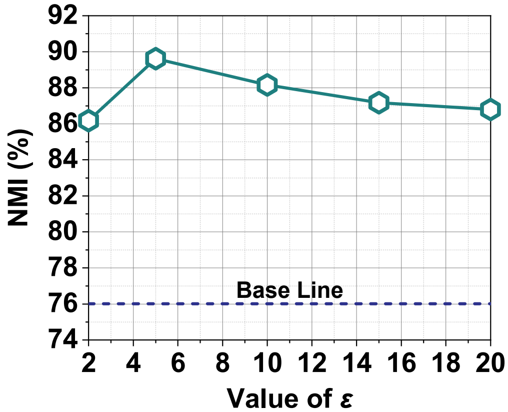
(a) NMI (b) Purity

Fig. S.1. Effects of ϑ on D1.

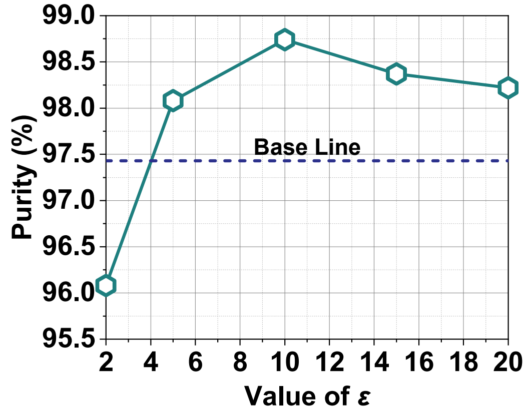


(a) NMI (b) Purity

Fig. S.2. Effects of λ on D1.

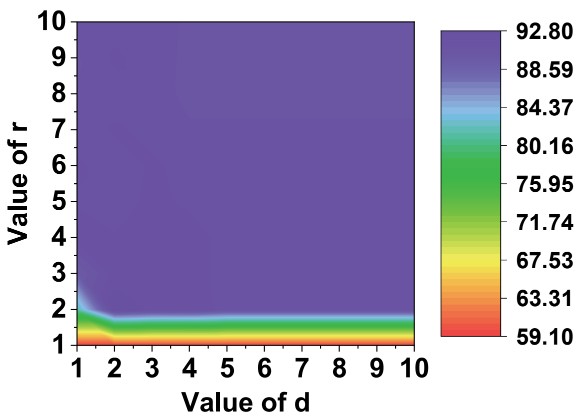


(a) NMI

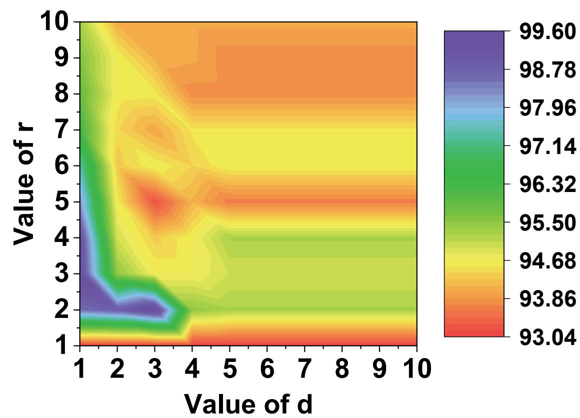


(b) Purity

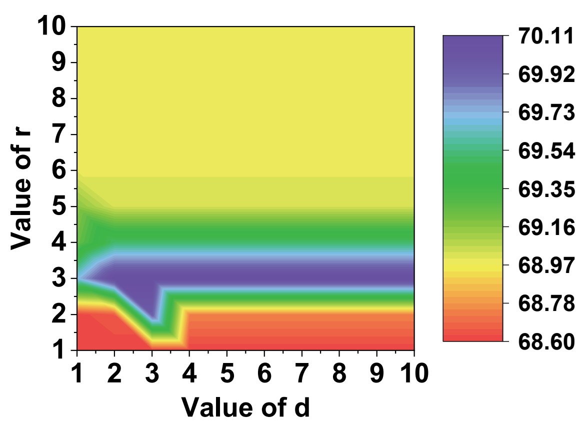
Fig. S.3. Effects of ε on D1.



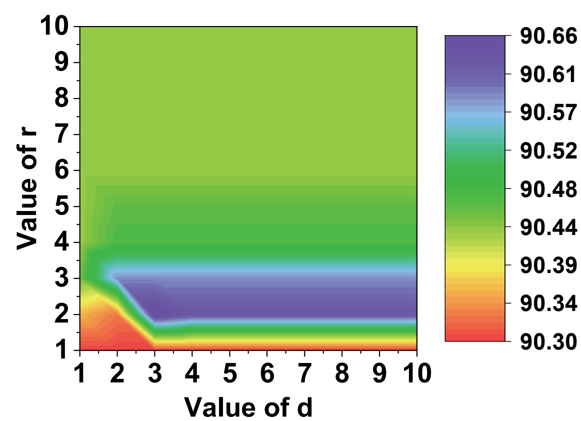
(a) NMI on D1



(b) Purity on D1

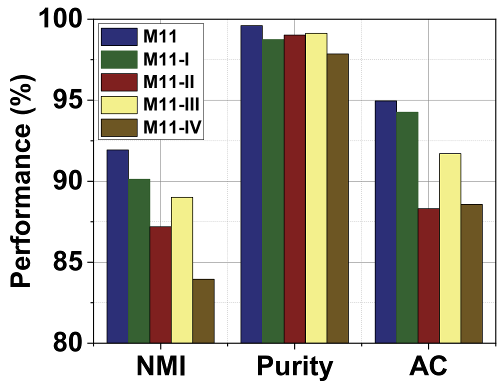


(c) NMI on D8

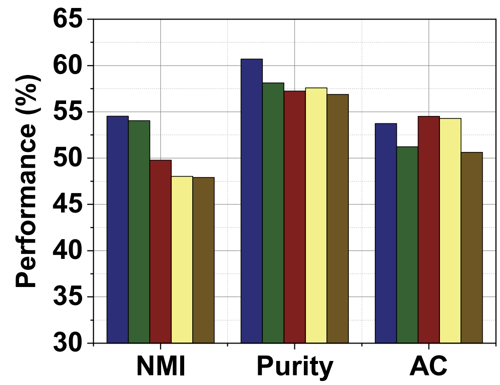


(d) Purity on D8

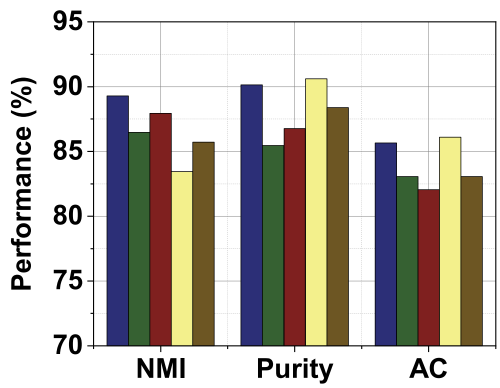
Fig. S.4. Effects of r and d on D1 and D8.



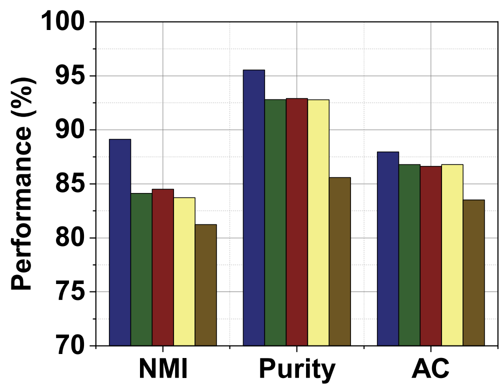
(a) D1



(b) D2



(c) D3



(d) D4

Fig. S.5. Performance comparison among HOP-NMF and its variants on D1-D4.

TABLE S.III
COMMUNITY DETECTION PERFORMANCE (NMI%±STD%) OF TESTED MODELS ON EACH NETWORK, INCLUDING WIN/LOSS COUNTS (✪, ○, ●, AND ⊛ INDICATE THAT M11-M14 WIN THE COMPARED MODELS).

Datasets / Models	D1	D2	D3	D4	D5	D6	D7	D8	✪Win/ Loss	○Win/ Loss	●Win/ Loss	⊛Win/ Loss
M1	42.21±1.46 ✪○●⊛	17.26±2.58 ✪○●⊛	68.08±6.66 ✪○●⊛	29.53±2.13 ✪○●⊛	23.47±3.87 ✪○●⊛	9.02±3.64 ✪○●⊛	35.65±0.40 ✪○●⊛	53.30±2.82 ✪○●⊛	8/0	8/0	8/0	8/0
M2	45.48±1.72 ✪○●⊛	16.95±2.10 ✪○●⊛	67.64±5.53 ✪○●⊛	29.39±6.36 ✪○●⊛	33.78±4.49 ✪○●⊛	12.51±0.49 ✪○●⊛	34.71±2.03 ✪○●⊛	62.44±3.39 ✪○●⊛	8/0	8/0	8/0	8/0
M3	46.55±1.87 ✪○●⊛	18.44±1.44 ✪○●⊛	76.14±5.13 ✪○●⊛	31.25±4.93 ✪○●⊛	19.79±5.19 ✪○●⊛	6.25±3.51 ✪○●⊛	34.08±2.27 ✪○●⊛	42.32±2.64 ✪○●⊛	8/0	8/0	8/0	8/0
M4	48.30±0.54 ✪○●⊛	18.59±1.85 ✪○●⊛	78.40±4.49 ✪○●⊛	33.75±7.20 ✪○●⊛	42.22±2.81 ✪○●⊛	15.60±2.52 ✪○●⊛	35.40±0.48 ✪○●⊛	63.65±0.89 ✪○●⊛	8/0	8/0	8/0	8/0
M5	48.23±1.65 ✪○●⊛	17.84±1.39 ✪○●⊛	70.31±1.95 ✪○●⊛	43.31±2.36 ✪○●⊛	42.33±1.59 ✪○●⊛	4.75±0.16 ✪○●⊛	39.66±0.50 ○●	36.18±0.69 ✪○●⊛	7/1	8/0	8/0	7/1
M6	68.20±1.20 ✪○●⊛	49.45±2.18 ✪○●⊛	62.64±6.95 ✪○●⊛	59.18±4.02 ✪○●⊛	53.94±7.03 ✪○●⊛	20.13±0.80 ✪○●⊛	44.23±4.00 ○	55.08±3.33 ✪○●⊛	7/1	8/0	7/1	7/1
M7	62.83±2.25 ✪○●⊛	38.99±2.47 ✪○●⊛	45.17±2.80 ✪○●⊛	71.92±3.88 ✪○●⊛	65.77±1.91 ✪○●	18.05±1.66 ✪○●⊛	41.32±2.00 ○●	63.53±1.49 ✪○●⊛	7/1	8/0	8/0	6/2
M8	75.31±2.11 ✪○●⊛	50.04±2.54 ✪○●⊛	80.55±3.55 ✪○●⊛	47.99±8.32 ✪○●⊛	68.03±4.51 ✪○●	16.40±2.77 ✪○●⊛	41.56±2.79 ○●	64.14±3.43 ✪○●⊛	7/1	8/0	8/0	6/2
M9	84.04±1.36 ✪●	50.63±1.84 ✪○●	84.91±3.80 ✪●	76.60±3.19 ✪○●⊛	73.44±0.71 ✪●	15.82±3.36 ✪○●⊛	**49.62±4.05** --	64.51±1.58 ✪○●⊛	7/1	4/4	7/1	3/5
M10	83.25±0.62 ✪●	51.56±1.77 ✪●	87.55±3.24 ✪●	75.75±0.06 ✪○●⊛	74.80±0.90 ✪●	22.65±2.63 ✪○⊛	47.51±1.44 ○	67.32±2.67 ✪○●⊛	7/1	4/4	6/2	3/5
M11✪	**91.94±0.23**	**54.52±0.93**	89.29±0.23	**89.12±10.06**	**84.31±0.08**	**29.92±2.02**	38.57±1.05	**69.94±1.82**	--	--	--	--
M12○	76.01±1.52	50.65±3.80	82.98±4.79	77.55±7.08	71.46±2.28	24.40±2.44	49.39±6.32	68.60±3.60	--	--	--	--
M13●	85.39±1.66	51.69±1.23	**93.52±0.11**	85.44±7.55	79.80±3.14	22.48±4.92	44.00±3.78	69.48±4.09	--	--	--	--
M14⊛	78.82±1.51	50.53±1.88	82.00±2.89	83.30±0.01	60.74±0.57	23.58±3.32	36.04±0.85	67.80±3.79	--	--	--	--

* *A lower Friedman rank value indicates a higher community detection accuracy.*

TABLE S.IV

COMMUNITY DETECTION PERFORMANCE (PURITY%±STD%) OF TESTED MODELS ON EACH NETWORK, INCLUDING WIN/LOSS COUNTS (✪, ○, ●, AND ⊛ INDICATE THAT M11-M14 WIN THE COMPARED MODELS).

Datasets / Models	D1	D2	D3	D4	D5	D6	D7	D8	✪Win/ Loss	○Win/ Loss	●Win/ Loss	⊛Win/ Loss
M1	93.30±1.36 ✪○●⊛	53.09±1.56 ✪○●⊛	81.20±0.02 ✪○●⊛	57.20±0.01 ✪○●⊛	64.54±0.83 ✪○●⊛	48.17±3.43 ✪○●⊛	91.52±0.27 ✪○●⊛	82.50±1.74 ✪○●⊛	8/0	8/0	8/0	8/0
M2	93.56±1.09 ✪○●⊛	59.32±7.12 ✪○●	78.86±3.19 ✪○●⊛	60.86±3.75 ✪○●⊛	66.61±5.29 ✪○●⊛	53.16±1.50 ✪○●⊛	91.50±2.17 ✪○●⊛	86.64±1.85 ✪○●⊛	8/0	8/0	8/0	7/1
M3	93.71±1.01 ✪○●⊛	58.61±4.97 ✪○●⊛	85.44±3.76 ✪○●⊛	52.27±5.06 ✪○●⊛	71.87±2.18 ✪○●⊛	46.87±1.21 ✪○●⊛	90.33±2.35 ✪○●⊛	76.98±1.56 ✪○●⊛	8/0	8/0	8/0	8/0
M4	94.33±0.58 ✪○●⊛	58.84±0.16 ✪○●⊛	91.82±0.08 ●	67.35±8.63 ✪○●⊛	71.21±2.84 ✪○●⊛	52.86±1.27 ✪○●⊛	91.10±0.13 ✪○●⊛	86.21±0.61 ✪○●⊛	7/1	7/1	8/0	7/1
M5	95.07±1.27 ✪○●⊛	58.90±0.43 ✪○●⊛	84.97±2.62 ✪○●⊛	76.66±3.02 ✪○●⊛	71.38±2.18 ✪○●⊛	27.81±0.21 ✪○●⊛	90.40±1.73 ✪○●⊛	26.33±0.99 ✪○●⊛	8/0	8/0	8/0	8/0
M6	96.75±0.34 ✪○●⊛	57.37±1.07 ✪○●⊛	89.51±3.30 ✪●	58.19±0.01 ✪○●⊛	71.31±1.64 ✪○●⊛	58.30±1.12 ✪⊛	92.93±0.49 ✪○●	83.32±1.15 ✪○●⊛	8/0	6/2	7/1	6/2
M7	95.82±1.31 ✪○●⊛	57.13±1.21 ✪○●⊛	84.74±1.87 ✪○●⊛	78.93±0.77 ✪●⊛	70.81±4.10 ✪○●⊛	53.42±0.84 ✪○●⊛	91.69±2.14 ✪○●⊛	87.11±1.57 ✪○●⊛	8/0	7/1	8/0	8/0
M8	94.94±1.18 ✪○●⊛	60.10±0.70 ✪○●	87.78±2.34 ✪●	65.25±9.92 ✪○●⊛	74.67±3.73 ✪●⊛	53.60±1.78 ✪○●⊛	94.40±0.33 ✪○	87.46±1.32 ✪○●⊛	8/0	6/2	7/1	5/3
M9	97.86±0.87 ✪●⊛	60.53±0.09 ✪	89.59±2.97 ✪●	87.11±2.76 ✪●	83.58±1.65 ✪●⊛	60.30±4.71 ✪	94.36±0.23 ✪○	88.19±0.42 ✪○●⊛	8/0	2/6	5/3	3/5
M10	97.60±0.28 ✪●⊛	59.93±1.12 ✪○●	90.06±2.39 ✪●	86.21±0.00 ✪●⊛	84.77±2.03 ✪⊛	60.33±3.17 ✪	94.93±0.40 ✪	88.72±1.06 ✪○●⊛	8/0	2/6	5/3	4/4
M11✪	**99.60±0.06**	**60.71±0.14**	90.13±1.00	**95.54±5.26**	**89.05±0.76**	**64.45±3.04**	**95.01±0.42**	**90.66±1.64**	--	--	--	--
M12○	97.43±0.89	60.48±0.29	87.27±2.18	77.61±1.24	73.45±2.11	54.45±2.45	94.52±0.11	89.31±1.85	--	--	--	--
M13●	98.53±0.17	60.22±0.53	**97.44±0.09**	87.28±4.25	84.53±0.33	57.34±3.67	93.30±1.89	90.52±1.83	--	--	--	--
M14⊛	98.15±0.51	59.22±1.01	87.04±6.06	86.53±0.01	86.69±2.02	58.35±3.79	92.04±0.19	89.58±2.00	--	--	--	--

** A lower Friedman rank value indicates a higher community detection accuracy.*

TABLE S.V

COMMUNITY DETECTION PERFORMANCE (AC%±STD%) OF TESTED MODELS ON EACH NETWORK, INCLUDING WIN/LOSS COUNTS (✪, ○, ●, AND ⊛ INDICATE THAT M11-M14 WIN THE COMPARED MODELS).

Datasets / Models	D1	D2	D3	D4	D5	D6	D7	D8	✪Win/ Loss	○Win/ Loss	●Win/ Loss	⊛Win/ Loss
M1	68.68±3.01 ✪○●⊛	43.64±0.38 ✪○●⊛	77.33±6.27 ✪○●⊛	48.18±0.90 ✪○●⊛	58.42±3.04 ✪○●⊛	40.95±4.35 ✪○●⊛	64.09±2.98 ✪○●⊛	67.46±2.29 ✪○●⊛	8/0	8/0	8/0	8/0
M2	72.14±2.49 ✪○●⊛	45.75±0.45 ✪○●⊛	73.51±5.49 ✪○●⊛	51.51±3.81 ✪○●⊛	63.16±0.93 ✪○●⊛	43.25±1.47 ✪○●⊛	61.79±6.19 ✪○●⊛	75.79±2.86 ✪○●⊛	8/0	8/0	8/0	8/0
M3	67.27±2.52 ✪○●⊛	45.51±2.81 ✪○●⊛	78.67±1.83 ✪○●⊛	49.28±0.76 ✪○●⊛	43.53±5.98 ✪○●⊛	39.24±3.80 ✪○●⊛	63.43±4.17 ✪○●⊛	66.16±0.81 ✪○●⊛	8/0	8/0	8/0	8/0
M4	76.21±0.40 ✪○●⊛	46.96±2.11 ✪○●⊛	83.60±2.39 ✪○●⊛	59.10±8.21 ✪○●⊛	63.49±0.55 ✪○●⊛	41.58±3.56 ✪○●⊛	62.30±3.15 ✪○●⊛	80.89±1.32 ✪○●⊛	8/0	8/0	8/0	8/0
M5	49.14±1.92 ✪○●⊛	46.72±1.99 ✪○●⊛	73.34±0.37 ✪○●⊛	45.73±4.06 ✪○●⊛	43.17±1.07 ✪○●⊛	41.41±4.16 ✪○●⊛	64.95±0.94 ✪○●⊛	81.44±1.42 ✪○●⊛	8/0	8/0	8/0	8/0
M6	76.62±1.94 ✪○●⊛	44.82±4.41 ✪○●⊛	**88.85±5.56** --	78.97±10.21 ✪○●⊛	75.22±2.97 ✪●	44.73±14.52 ✪○●⊛	66.32±0.36 ✪○●⊛	72.56±1.27 ✪○●⊛	7/1	6/2	7/1	6/2
M7	75.52±2.53 ✪○●⊛	46.36±2.53 ✪○●⊛	85.42±7.22 ✪○	72.07±9.71 ✪○●⊛	69.77±2.39 ✪○●⊛	51.40±3.62 ✪	62.06±6.11 ✪○●⊛	83.12±3.66 ✪	8/0	6/2	5/3	5/3
M8	79.96±0.98 ✪○●⊛	47.81±2.18 ✪○●⊛	81.67±6.09 ✪○●⊛	63.33±6.64 ✪○●⊛	72.65±6.69 ✪○●	43.78±1.45 ✪○●⊛	86.77±2.41 ✪	81.61±1.43 ✪○●⊛	8/0	7/1	7/1	6/2
M9	86.76±0.42 ✪●	44.87±4.56 ✪○●⊛	80.11±2.09 ✪○●⊛	86.28±9.13 ✪●	73.28±1.23 ✪○●	47.49±2.53 ✪○●⊛	89.54±3.35 ✪	81.75±1.03 ✪○⊛	8/0	5/3	6/2	4/4
M10	88.13±1.40 ✪●	48.58±1.47 ✪○●⊛	85.13±3.43 ✪○●	83.64±8.17 ✪●⊛	76.05±3.35 ✪●	51.12±3.08 ✪●⊛	89.34±4.40 ✪	81.47±2.13 ✪○●⊛	8/0	3/5	7/1	4/4
M11✪	**94.95±0.28**	**53.74±0.33**	85.66±5.11	**87.95±5.69**	**83.65±1.04**	**51.56±1.05**	**90.44±0.98**	**84.27±3.43**	--	--	--	--
M12○	82.93±0.96	52.02±0.13	86.11±1.63	79.95±9.52	74.26±1.38	50.81±1.19	81.38±7.32	82.54±2.46	--	--	--	--
M13●	94.27±0.52	53.15±1.38	85.23±3.75	86.78±8.92	81.47±1.96	51.26±4.29	85.01±6.54	81.69±1.74	--	--	--	--
M14⊛	86.28±0.60	49.51±1.91	83.91±1.23	86.03±0.02	71.75±4.94	51.16±1.31	86.28±1.59	82.04±2.50	--	--	--	--

** A lower Friedman rank value indicates a higher community detection accuracy.*

TABLE S.VI

AVERAGE FRIEDMAN RANKS OF ALL TESTED MODELS BY TAKING ALL THE TESTING CASES IN TABLES S.III-V INTO ACCOUNT.

Models	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**	**M8**	**M9**	**M10**	**M11**	**M12**	**M13**	**M14**
Ranks*	12.50	11.58	12.17	9.71	11.25	8.25	8.46	7.17	4.96	4.21	**1.54**	4.75	3.17	5.29

** A lower Friedman rank value indicates a higher community detection accuracy.*

TABLE S.VII

RESULTS OF WILCOXON SIGNED-RANK TEST BY TAKING INTO ACCOUNT ALL OF THE TESTING CASES IN TABLES S.III-V.

VS.		**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**	**M8**	**M9**	**M10**
	R+	300	300	300	299	299	289	296	295	278	279
M11	*R-*	0	0	0	1	1	11	4	5	22	21
	p-value	**9.7112E-06**	**9.7112E-06**	**9.7112E-06**	**1.1035E-05**	**1.1035E-05**	**3.7926E-05**	**1.6113E-05**	**1.8234E-05**	**1.3481E-04**	**1.2060E-04**
	R+	300	300	300	291	300	273	292	271	120	84
M12	*R-*	0	0	0	9	0	27	8	29	180	216
	p-value	**9.7112E-06**	**9.7112E-06**	**9.7112E-06**	**2.9787E-06**	**9.7112E-06**	**2.3263E-04**	**2.6402E-05**	**2.8780E-04**	0.8082	0.9713
	R+	300	300	300	291	300	288	294	292	248.5	233
M13	*R-*	0	0	0	9	0	12	6	8	51.5	67
	p-value	**9.7112E-06**	**9.7112E-06**	**9.7112E-06**	**2.9787E-06**	**9.7112E-06**	**4.2726E-05**	**2.0639E-05**	**2.6379E-05**	**0.0026**	**0.0092**
	R+	300	299	300	289	293	259	265	236.5	142	89
M14	*R-*	0	1	0	11	7	41	35	63.5	158	211
	p-value	**9.7112E-06**	**1.1035E-05**	**9.7112E-06**	**3.7926E-05**	**2.3363E-05**	**9.6760E-04**	**5.3503E-04**	**0.0070**	0.5959	0.9606

** The accepted hypotheses with a significance level of 0.05 are highlighted.*