

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 symbols and acronyms 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. SYMBOLS AND ACRONYMS

Main acronyms and symbols of the paper are summarized in Tables SI and SII.

TABLE SI
MAIN ACRONYM APPOINTMENT

Acronym	Description
IoT	Internet of Things
SNMF	Symmetric 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

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:

TABLE SII
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 .
$U^{m \times K}$	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 .
$\phi_{ik}, \psi_{jk}, \kappa_{sk}$	Single entries in Φ, Ψ and K .
ζ, 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.
\emptyset	Null set.
$ \cdot $	Cardinality of an enclosed set.
$\ \cdot\ _F$	Frobenius norm of an enclosed matrix
$\text{Tr}(\cdot)$	Trace of an enclosed matrix.

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

which results in

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

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:

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

which results in

$$(Y^{(*)}U^T U + Y^{(*)} - \tilde{A}U - X)_{jk} y_{jk}^{(*)} = 0, \quad (\text{S4})$$

which 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:

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

which results in

$$(U^{(*)}Y^T Y + U^{(*)} - \tilde{A}^T Y - X)_{sk} u_{sk}^{(*)} = 0, \quad (\text{S6})$$

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:

$$x^{(t+1)} = \arg \min_x \Gamma(x, x^{(t)}). \quad (\text{S7})$$

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.,

$$F'(x_{ik}) = \frac{\partial J_{CGF}}{\partial x_{ik}} = (-\mathcal{G}Y - \mathcal{G}U + XX^T X - \tilde{A}X + \lambda \tilde{L}X)_{ik} + 2\mathcal{G}x_{ik}, \quad (\text{S8})$$

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

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

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

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:

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

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:

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

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

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

It is easy to show that

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

Similarly, we have

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

and

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

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

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

On the other hand, we have

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

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 $F(x_{ik})$. \square

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).

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.

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:

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

which is identical with the learning rule in (23a). Note that x_{ik} is an arbitrary entry of X . Hence, $\forall i \in \{1, 2, \dots, n\}, k \in \{1, 2, \dots, K\}$, $F(x_{ik})$ 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.,

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

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

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

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

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:

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

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:

$$(\mathcal{G}(YU^T U)_{jk} + \mathcal{G}y_{jk}^{(t)}) / y_{jk}^{(t)} \geq \mathcal{G}(U^T U)_{kk} + \mathcal{G}. \tag{S24}$$

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

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

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

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

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:

$$\begin{aligned} y_{jk}^{(t+1)} &= \arg \min_{y_{jk}} \Gamma(y_{jk}, y_{jk}^{(t)}) \\ \Rightarrow F'(y_{jk}^{(t)}) + \frac{\mathcal{G}(YU^T U)_{jk} + \mathcal{G}y_{jk}^{(t)}}{y_{jk}^{(t)}} (y_{jk} - y_{jk}^{(t)}) &= 0 \\ \Rightarrow y_{jk} &= y_{jk}^{(t)} - y_{jk}^{(t)} \frac{F'(y_{jk}^{(t)})}{\mathcal{G}(YU^T U)_{jk} + \mathcal{G}y_{jk}^{(t)}} \Rightarrow y_{jk}^{(t+1)} = y_{jk}^{(t)} \frac{(\tilde{A}U + X)_{jk}}{(YU^T U + Y)_{jk}}, \end{aligned} \quad (S27)$$

which is identical with the learning rule in (23b). Similarly, $\forall j \in \{1, 2, \dots, n\}, k \in \{1, 2, \dots, K\}$, $F(y_{jk})$ 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.,

$$F'(u_{sk}) = \frac{\partial J_{CGF}}{\partial u_{sk}} = (\mathcal{G}UY^T Y - \mathcal{G}\tilde{A}^T Y - \mathcal{G}X)_{sk} + \mathcal{G}u_{sk}, \quad (S28)$$

$$F''(u_{sk}) = \frac{\partial^2 J_{CGF}}{\partial u_{sk}^2} = \mathcal{G}(Y^T Y)_{kk} + \mathcal{G}. \quad (S29)$$

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 $F(u_{sk})$.

$$\Gamma(u_{sk}, u_{sk}^{(t)}) = F(u_{sk}^{(t)}) + F'(u_{sk}^{(t)})(u_{sk} - u_{sk}^{(t)}) + \frac{\mathcal{G}(UY^T Y)_{sk} + \mathcal{G}u_{sk}^{(t)}}{2u_{sk}^{(t)}} (u_{sk} - u_{sk}^{(t)})^2. \quad (S30)$$

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.,

$$F(u_{sk}) = F(u_{sk}^{(t)}) + F'(u_{sk}^{(t)})(u_{sk} - u_{sk}^{(t)}) + \frac{1}{2}(\mathcal{G}(Y^T Y)_{kk} + \mathcal{G})(u_{sk} - u_{sk}^{(t)})^2. \quad (S31)$$

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:

$$(\mathcal{G}(UY^T Y)_{sk} + \mathcal{G}u_{sk}^{(t)})/u_{sk}^{(t)} \geq \mathcal{G}(Y^T Y)_{kk} + \mathcal{G}. \quad (S32)$$

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

$$(UY^T Y)_{sk} \geq u_{sk}^{(t)}(Y^T Y)_{kk}. \quad (S33)$$

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

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

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:

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

which is identical with the learning rule in (23c). Similarly, $\forall s \in \{1, 2, \dots, n\}$, $k \in \{1, 2, \dots, K\}$, $F(u_{sk})$ 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.

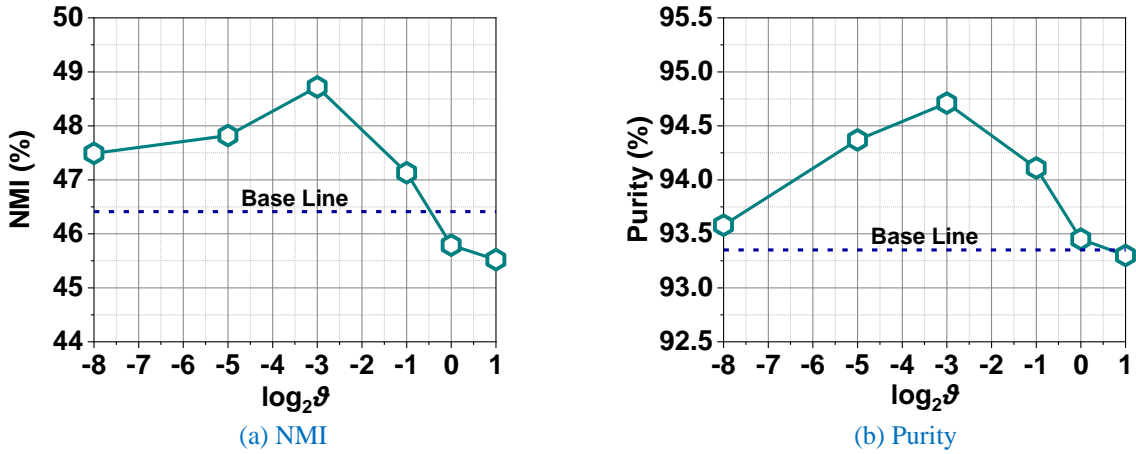


Fig. S1. Effects of ρ on D1.

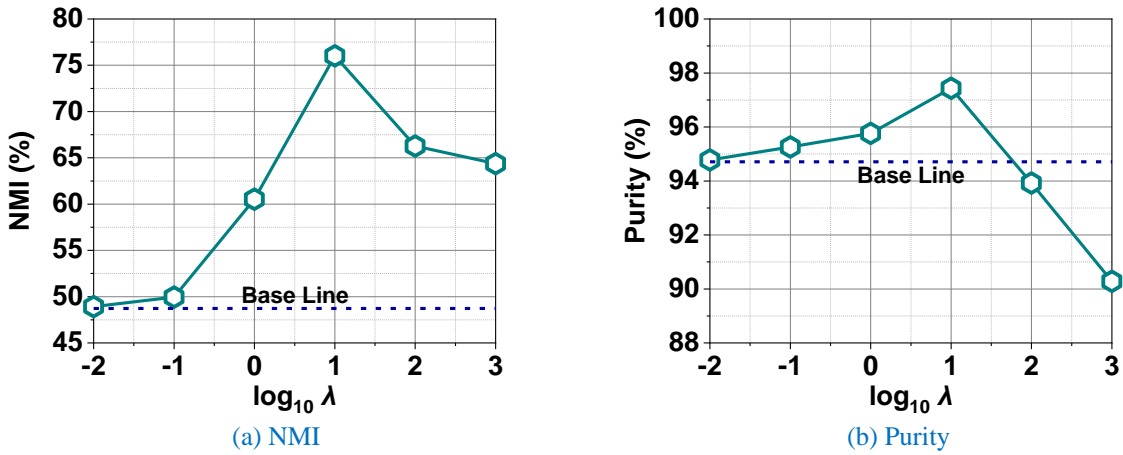


Fig. S2. Effects of λ on D1.

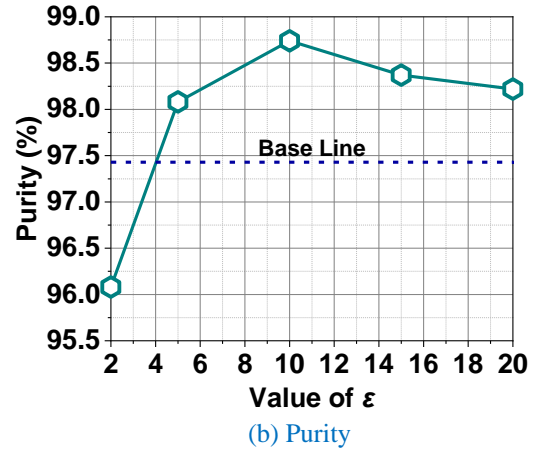
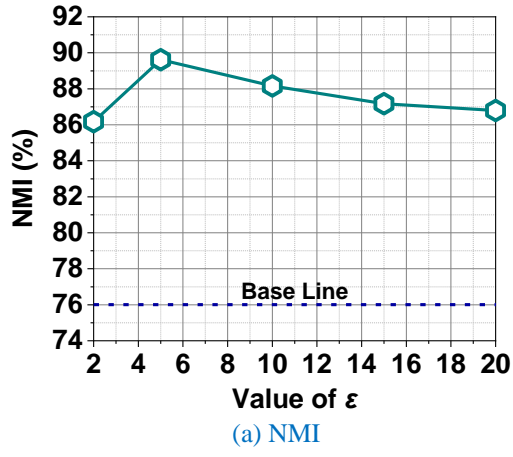


Fig. S3. Effects of ε on D1.

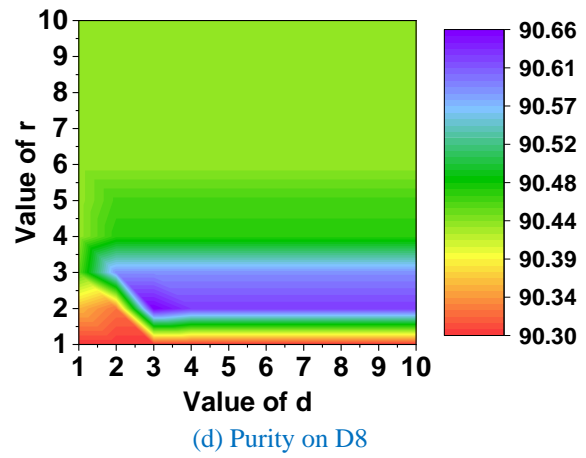
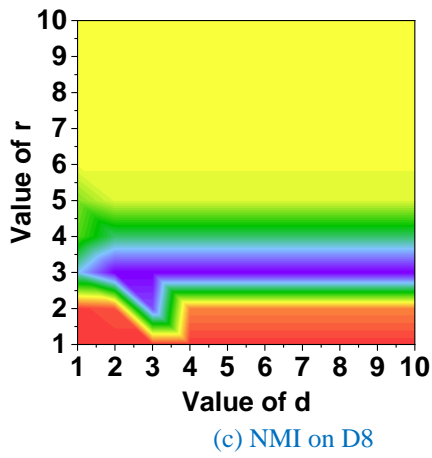
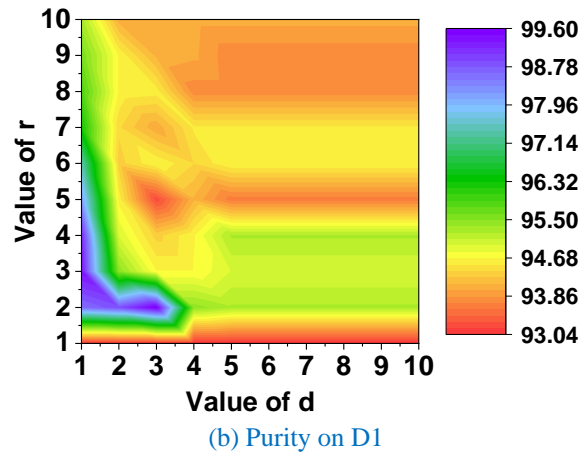
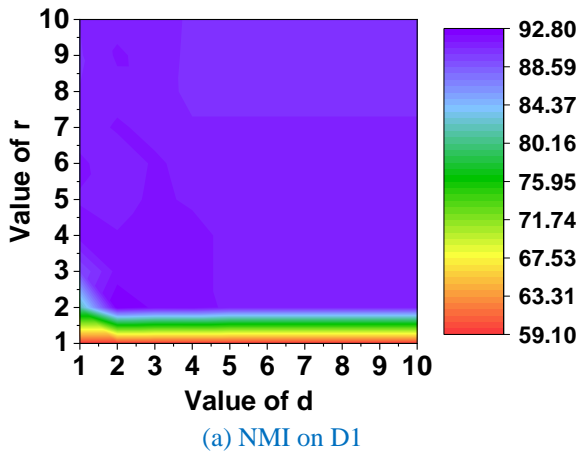
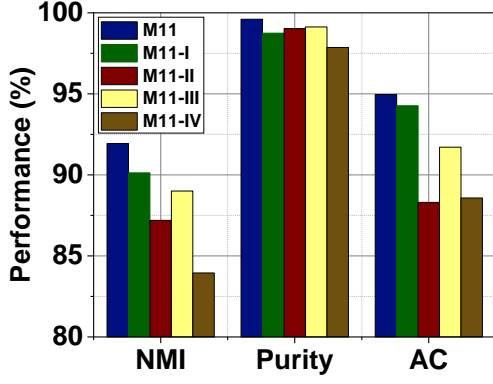
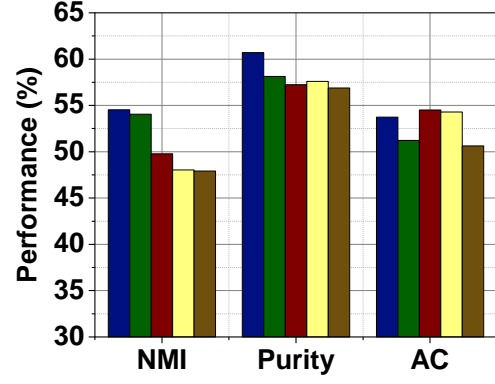


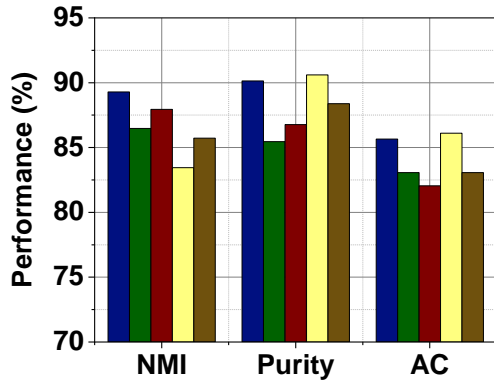
Fig. S4. Effects of r and d on D1 and D8.



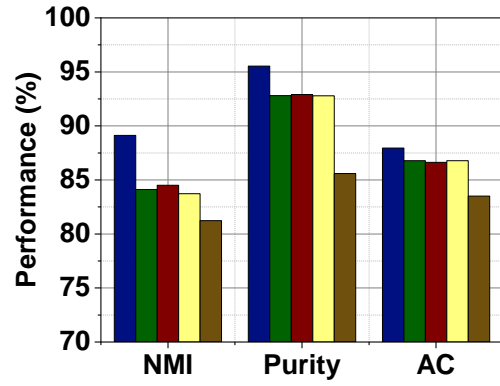
(a) D1



(b) D2



(c) D3



(d) D4

Fig. S5. Performance comparison among HOP-NMF and its variants on D1-D4.

TABLE III

COMMUNITY DETECTION PERFORMANCE (NMI%±STD%) OF TESTED MODELS ON EACH NETWORK, INCLUDING WIN/LOSS COUNTS AND FRIEDMAN TEST (●, ○, ●, ●, 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	Friedman Ranks*
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	12.38
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	12.38
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	12.25
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	10.00
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	11.13
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	8.38
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	8.63
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	7.13
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	4.63
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	3.88
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	--	--	--	--	2.13
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	--	--	--	--	3.88
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	--	--	--	--	2.63
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	--	--	--	--	5.63

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

TABLE SIV

COMMUNITY DETECTION PERFORMANCE (PURITY% \pm STD%) OF TESTED MODELS ON EACH NETWORK, INCLUDING WIN/LOSS COUNTS AND FRIEDMAN TEST (●, ○, ●, AND ⊙ INDICATE THAT M11-M14 WIN THE COMPARED MODELS).

Datasets	D1	D2	D3	D4	D5	D6	D7	D8	●Win/ Loss	○Win/ Loss	●Win/ Loss	⊙Win/ Loss	Friedman Ranks*
Models													
M1	93.30 \pm 1.36 ●○●●	53.09 \pm 1.56 ●●●●	81.20 \pm 0.02 ●●●●	57.20 \pm 0.01 ●●●●	64.54 \pm 0.83 ●●●●	48.17 \pm 3.43 ●●●●	91.52 \pm 0.27 ●○●●	82.50 \pm 1.74 ●○●●	8/0	8/0	8/0	8/0	12.75
M2	93.56 \pm 1.09 ●○●●	59.32 \pm 7.12 ●○●●	78.86 \pm 3.19 ●○●●	60.86 \pm 3.75 ●○●●	66.61 \pm 5.29 ●○●●	53.16 \pm 1.50 ●○●●	91.50 \pm 2.17 ●○●●	86.64 \pm 1.85 ●○●●	8/0	8/0	8/0	7/1	11.00
M3	93.71 \pm 1.01 ●○●●	58.61 \pm 4.97 ●○●●	85.44 \pm 3.76 ●○●●	52.27 \pm 5.06 ●○●●	71.87 \pm 2.18 ●○●●	46.87 \pm 1.21 ●○●●	90.33 \pm 2.35 ●○●●	76.98 \pm 1.56 ●○●●	8/0	8/0	8/0	8/0	11.88
M4	94.33 \pm 0.58 ●○●●	58.84 \pm 0.16 ●○●●	91.82 \pm 0.08 ●○●●	67.35 \pm 8.63 ●○●●	71.21 \pm 2.84 ●○●●	52.86 \pm 1.27 ●○●●	91.10 \pm 0.13 ●○●●	86.21 \pm 0.61 ●○●●	7/1	7/1	8/0	7/1	9.50
M5	95.07 \pm 1.27 ●○●●	58.90 \pm 0.43 ●○●●	84.97 \pm 2.62 ●○●●	76.66 \pm 3.02 ●○●●	71.38 \pm 2.18 ●○●●	27.81 \pm 0.21 ●○●●	90.40 \pm 1.73 ●○●●	26.33 \pm 0.99 ●○●●	8/0	8/0	8/0	8/0	10.88
M6	96.75 \pm 0.34 ●○●●	57.37 \pm 1.07 ●○●●	89.51 \pm 3.30 ●○●●	58.19 \pm 0.01 ●○●●	71.31 \pm 1.64 ●○●●	58.30 \pm 1.12 ●○●●	92.93 \pm 0.49 ●○●●	83.32 \pm 1.15 ●○●●	8/0	6/2	7/1	6/2	8.75
M7	95.82 \pm 1.31 ●○●●	57.13 \pm 1.21 ●○●●	84.74 \pm 1.87 ●○●●	78.93 \pm 0.77 ●○●●	70.81 \pm 4.10 ●○●●	53.42 \pm 0.84 ●○●●	91.69 \pm 2.14 ●○●●	87.11 \pm 1.57 ●○●●	8/0	7/1	8/0	8/0	9.63
M8	94.94 \pm 1.18 ●○●●	60.10 \pm 0.70 ●○●●	87.78 \pm 2.34 ●○●●	65.25 \pm 9.92 ●○●●	74.67 \pm 3.73 ●○●●	53.60 \pm 1.78 ●○●●	94.40 \pm 0.33 ●○●●	87.46 \pm 1.32 ●○●●	8/0	6/2	7/1	5/3	7.13
M9	97.86 \pm 0.87 ●○●●	60.53 \pm 0.09 ●○●●	89.59 \pm 2.97 ●○●●	87.11 \pm 2.76 ●○●●	83.58 \pm 1.65 ●○●●	60.30 \pm 4.71 ●○●●	94.36 \pm 0.23 ●○●●	88.19 \pm 0.42 ●○●●	8/0	2/6	5/3	3/5	4.13
M10	97.60 \pm 0.28 ●○●●	59.93 \pm 1.12 ●○●●	90.06 \pm 2.39 ●○●●	86.21 \pm 0.00 ●○●●	84.77 \pm 2.03 ●○●●	60.33 \pm 3.17 ●○●●	94.93 \pm 0.40 ●○●●	88.72 \pm 1.06 ●○●●	8/0	2/6	5/3	4/4	4.00
M11●	99.60\pm0.06	60.71\pm0.14	90.13 \pm 1.00	95.54\pm5.26	89.05\pm0.76	64.45\pm3.04	95.01\pm0.42	90.66\pm1.64	--	--	--	--	1.25
M12○	97.43 \pm 0.89	60.48 \pm 0.29	87.27 \pm 2.18	77.61 \pm 1.24	73.45 \pm 2.11	54.45 \pm 2.45	94.52 \pm 0.11	89.31 \pm 1.85	--	--	--	--	5.63
M13●	98.53 \pm 0.17	60.22 \pm 0.53	97.44\pm0.09	87.28 \pm 4.25	84.53 \pm 0.33	57.34 \pm 3.67	93.30 \pm 1.89	90.52 \pm 1.83	--	--	--	--	3.38
M14⊙	98.15 \pm 0.51	59.22 \pm 1.01	87.04 \pm 6.06	86.53 \pm 0.01	86.69 \pm 2.02	58.35 \pm 3.79	92.04 \pm 0.19	89.58 \pm 2.00	--	--	--	--	5.13

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

TABLE SV

COMMUNITY DETECTION PERFORMANCE (AC% \pm STD%) OF TESTED MODELS ON EACH NETWORK, INCLUDING WIN/LOSS COUNTS AND FRIEDMAN TEST (●, ○, ●, AND ⊙ INDICATE THAT M11-M14 WIN THE COMPARED MODELS).

Datasets	D1	D2	D3	D4	D5	D6	D7	D8	●Win/ Loss	○Win/ Loss	●Win/ Loss	⊙Win/ Loss	Friedman Ranks*
Models													
M1	68.68 \pm 3.01 ●○●●	43.64 \pm 0.38 ●○●●	77.33 \pm 6.27 ●○●●	48.18 \pm 0.90 ●○●●	58.42 \pm 3.04 ●○●●	40.95 \pm 4.35 ●○●●	64.09 \pm 2.98 ●○●●	67.46 \pm 2.29 ●○●●	8/0	8/0	8/0	8/0	12.38
M2	72.14 \pm 2.49 ●○●●	45.75 \pm 0.45 ●○●●	73.51 \pm 5.49 ●○●●	51.51 \pm 3.81 ●○●●	63.16 \pm 0.93 ●○●●	43.25 \pm 1.47 ●○●●	61.79 \pm 6.19 ●○●●	75.79 \pm 2.86 ●○●●	8/0	8/0	8/0	8/0	11.38
M3	67.27 \pm 2.52 ●○●●	45.51 \pm 2.81 ●○●●	78.67 \pm 1.83 ●○●●	49.28 \pm 0.76 ●○●●	43.53 \pm 5.98 ●○●●	39.24 \pm 3.80 ●○●●	63.43 \pm 4.17 ●○●●	66.16 \pm 0.81 ●○●●	8/0	8/0	8/0	8/0	12.38
M4	76.21 \pm 0.40 ●○●●	46.96 \pm 2.11 ●○●●	83.60 \pm 2.39 ●○●●	59.10 \pm 8.21 ●○●●	63.49 \pm 0.55 ●○●●	41.58 \pm 3.56 ●○●●	62.30 \pm 3.15 ●○●●	80.89 \pm 1.32 ●○●●	8/0	8/0	8/0	8/0	9.63
M5	49.14 \pm 1.92 ●○●●	46.72 \pm 1.99 ●○●●	73.34 \pm 0.37 ●○●●	45.73 \pm 4.06 ●○●●	43.17 \pm 1.07 ●○●●	41.41 \pm 4.16 ●○●●	64.95 \pm 0.94 ●○●●	81.44 \pm 1.42 ●○●●	8/0	8/0	8/0	8/0	11.75
M6	76.62 \pm 1.94 ●○●●	44.82 \pm 4.41 ●○●●	88.85\pm5.56	78.97 \pm 10.21 ●○●●	75.22 \pm 2.97 ●○●●	44.73 \pm 14.52 ●○●●	66.32 \pm 0.36 ●○●●	72.56 \pm 1.27 ●○●●	7/1	6/2	7/1	6/2	7.63
M7	75.52 \pm 2.53 ●○●●	46.36 \pm 2.53 ●○●●	85.42 \pm 7.22 ●○●●	72.07 \pm 9.71 ●○●●	69.77 \pm 2.39 ●○●●	51.40 \pm 3.62 ●○●●	62.06 \pm 6.11 ●○●●	83.12 \pm 3.66 ●○●●	8/0	6/2	5/3	5/3	7.13
M8	79.96 \pm 0.98 ●○●●	47.81 \pm 2.18 ●○●●	81.67 \pm 6.09 ●○●●	63.33 \pm 6.64 ●○●●	72.65 \pm 6.69 ●○●●	43.78 \pm 1.45 ●○●●	86.77 \pm 2.41 ●○●●	81.61 \pm 1.43 ●○●●	8/0	7/1	7/1	6/2	7.25
M9	86.76 \pm 0.42 ●○●●	44.87 \pm 4.56 ●○●●	80.11 \pm 2.09 ●○●●	86.28 \pm 9.13 ●○●●	73.28 \pm 1.23 ●○●●	47.49 \pm 2.53 ●○●●	89.54 \pm 3.35 ●○●●	81.75 \pm 1.03 ●○●●	8/0	5/3	6/2	4/4	6.13
M10	88.13 \pm 1.40 ●○●●	48.58 \pm 1.47 ●○●●	85.13 \pm 3.43 ●○●●	83.64 \pm 8.17 ●○●●	76.05 \pm 3.35 ●○●●	51.12 \pm 3.08 ●○●●	89.34 \pm 4.40 ●○●●	81.47 \pm 2.13 ●○●●	8/0	3/5	7/1	4/4	4.75
M11●	94.95\pm0.28	53.74\pm0.33	85.66 \pm 5.11	87.95\pm5.69	83.65\pm1.04	51.56\pm1.05	90.44\pm0.98	84.27\pm3.43	--	--	--	--	1.25
M12○	82.93 \pm 0.96	52.02 \pm 0.13	86.11 \pm 1.63	79.95 \pm 9.52	74.26 \pm 1.38	50.81 \pm 1.19	81.38 \pm 7.32	82.54 \pm 2.46	--	--	--	--	4.75
M13●	94.27 \pm 0.52	53.15 \pm 1.38	85.23 \pm 3.75	86.78 \pm 8.92	81.47 \pm 1.96	51.26 \pm 4.29	85.01 \pm 6.54	81.69 \pm 1.74	--	--	--	--	3.50
M14⊙	86.28 \pm 0.60	49.51 \pm 1.91	83.91 \pm 1.23	86.03 \pm 0.02	71.75 \pm 4.94	51.16 \pm 1.31	86.28 \pm 1.59	82.04 \pm 2.50	--	--	--	--	5.13

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

TABLE SVI

AVERAGE FRIEDMAN RANKS OF ALL TESTED MODELS BY TAKING INTO ACCOUNT ALL OF THE TESTING CASES IN TABLES V-VII AT THE SAME TIME.

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 SVII
RESULTS OF WILCOXON SIGNED-RANK TEST.

VS.		M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
M11	R+	300	300	300	299	299	289	296	295	278	279
	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
M12	R+	300	300	300	291	300	273	292	271	120	84
	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
M13	R+	300	300	300	291	300	288	294	292	248.5	233
	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
M14	R+	300	299	300	289	293	259	265	236.5	142	89
	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.