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

Zhigang Liu, 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 the proof of *Theorem* 1 and *Convergence Proof* in Section II.

## II. 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) in the manuscript 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:

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

which results in

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

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:

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

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:

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

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:

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

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

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

It is easy to show that

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

Similarly, we have

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

and

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

By combining (S15)-(S16), 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.

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

On the other hand, we have

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

Based on the inferences in (S17) and (S18), we conclude that (S13) holds when 0<$\beta$≤1/3, resulting in the establishment of (S12). Note that it is hard to rigorously show that (S13) still holds when 1/3<$\beta$≤1. But in fact, (S13) most likely would hold when 1/3<$\beta$≤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 $\Gamma(x_{ik}, x_{ik}^{(t)})\ge F(x_{ik})$, 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:

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

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 $\Gamma(y_{jk}, y_{jk}^{(t)})\ge F(y_{jk})$. 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 $\Gamma(y_{jk}, y_{jk}^{(t)})\ge F(y_{jk})$ is equivalent to the following condition:

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

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

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

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

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

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:

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

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.