3N4-OS-28b-3

TensorFlowを用いた半教師あり非負値行列分解に基づく制約付きコミュニティ抽出

Constrained community detection based on semi-supervised non-negative matrix factorization on TensorFlow

貫井 駿　　村田 剛志
Shun Nukui　Tsuyoshi Murata

東京工業大学 情報理工学院 情報工学系
Department of Comuper Science, School of Computing, Tokyo Institute of Technology

Community detection is an important task to understand network structures. Non-negative Matrix Factorization(NMF) is a useful method to discover the latent states of nodes or overlapping communties in the networks. NMF is basically an unsupervised learning method. However, in practice we often have prior knowledge about community memberships in a part of the network. The semi-supervised NMF method for community detection can be used to integrate prior knowledge into the NMF process. In this paper, we propose a novel method based on semi-supervised NMF. By combining the improved objective function and the gradient-based optimization method, we have improved the performance of existing semi-supervised NMF. We implemented our method on Tensorflow platform and performed experiments on real world networks. The empirical results show that our method outperform existing methods in term of accuracy.

1. はじめに

ネットワークはノードとエッジから成るデータ構造であり，物理学，生物学，情報工学など多くの分野において見受けられる．そのネットワークにおいてエッジが密な部分グラフをコミュニティという．ネットワークからコミュニティ構造を発見する手法をコミュニティ抽出と呼び，ネットワーク構造を理解するための重要なタスクである [Fortunato 10]．コミュニティ抽出ではモジュラリティを用いた手法 [Newman 04] がよく使われているが，最近ではノードを低次元の潜在ベクトルに変換してクラスタリングする手法が研究されている．その手法のひとつに非負値行列分解 (NMF) によるコミュニティ抽出がある．NMF によるコミュニティ抽出では隣接行列 $\mathbf{A}$ をより低次元の非負値行列 $\mathbf{W}, \mathbf{H}$ の積に分解し，各行ベクトルをクラスタリングする [Psorakis 11]．一般にコミュニティ抽出は教師なし学習に分類される技術であり，事前知識を与える必要がない．一方現実ではネットワークの一部についてコミュニティの事前知識を持っており, それを活用したい場面がしばしばある. 事前知識を制約として与えてコミュニティの抽出精度を高める手法を制約付きコミュニティ抽出と呼ぶ．半教師ありの NMF を用いたコミュニティ抽出は [Yang 15] で提案されているように，目的関数にペアワイズのコミュニティ制約項を導入し，それを最適化するという方法をとる．NMF は非負値制約から乗法更新規則を導出するのが一般的な解法であるが [Lee 00]，制約項を導入することにより目的関数の形が複雑になると更新規則を求めるのが困難になる．

そこで本研究では (1) 目的関数における制約項の改良と (2) 非負化処理をした Adaptive Moment Estimation(Adam)[Kingma 15] による勾配ベースの最適化という 2 つのアプローチによる新しい半教師あり NMF アルゴリズムを提案する．また 4 つの実ネットワークを用いたコミュニティ抽出実験において，TensorFlow で実装した提案手法が従来法よりも精度が大きく上回ることを示す．

連絡先: 貫井 駿，東京工業大学 情報理工学院 情報工学系 村田剛志研究室，東京都目黒区大岡山 2-12-1 W8-59, nukui.s.aa@m.titech.ac.jp

2. 関連研究

本節では本研究と関係のある先行研究を説明する．

2.1 非負値行列分解によるコミュニティ抽出

ネットワーク $G = (V, E)$ の隣接行列を $A \in \mathcal{R}_+^{N \times N}$, ただし $N = |V|$ とする．[Psorakis 11] で用いられたネットワーク生成モデルを仮定すると，ノード i,j 間のエッジの重み a_{ij} は潜在変数 $w_{ik} \in \mathcal{R}_+$, $h_{jk} \in \mathcal{R}_+$ を用いて式 (1) のように表される．

$$a_{ij} \simeq \sum_k w_{ik} h_{jk} \tag{1}$$

コミュニティの潜在因子の数を $K \in \mathcal{N}_+$ とすると，隣接行列 $\mathbf{A}$ は非負値行列 $\mathbf{W} \in \mathcal{R}_+^{N \times K}$, $\mathbf{H} \in \mathcal{R}_+^{N \times K}$ によって式 (2) のように表される．

$$\mathbf{A} \simeq \mathbf{W}\mathbf{H}^T \tag{2}$$

NMF では式 (2) の両辺の差分を最小化する $\mathbf{W}$,$\mathbf{H}$ を探索することが目的である．本論文では差分を表す損失関数として，式 (3) で表される Frobenius ノルムを用いる．

$$\mathcal{L}_{LSE} = ||\mathbf{A} - \mathbf{W}\mathbf{H}^T||_F^2 \tag{3}$$

式 (3) を最小化する $\mathbf{W}$,$\mathbf{H}$ を計算する最も一般的な方法は [Lee 00] により提案された乗法更新規則 (multiplicative update rules) と呼ばれる手法で，更新式 (4) を収束するまで繰り返し実行することで近似解を得る．

$$w_{ik} \leftarrow w_{ik} \frac{(\mathbf{A}\mathbf{H})_{ik}}{(\mathbf{W}\mathbf{H}^T\mathbf{H})_{ik}}, h_{jk} \leftarrow h_{jk} \frac{(\mathbf{A}^T\mathbf{W})_{jk}}{(\mathbf{H}\mathbf{W}^T\mathbf{H})_{jk}} \tag{4}$$

乗法更新規則は理論的に収束が証明されている．また [Lin 05] では Projected Gradient 法 (PG 法) と呼ばれる勾配法による最適化手法が提案されており，式 (5) により変数の更新を行う．

$$\begin{aligned} \mathbf{W} &\leftarrow max(0, \mathbf{W} - \alpha_t \nabla_{\mathbf{W}} \mathcal{L}_{LSE}) \\ \mathbf{H} &\leftarrow max(0, \mathbf{H} - \alpha_t \nabla_{\mathbf{H}} \mathcal{L}_{LSE}) \end{aligned} \tag{5}$$

式中の α_t は t 回目の更新時のステップサイズである．PG 法では $\mathbf{W}$,$\mathbf{H}$ の非負値制約を満たすため，勾配減算後に負値をとる場合に 0 へ写像する．それにより収束が安定しないという特徴がある．

ネットワークの隣接行列 $\mathbf{A}$ に対して式 (3) を最小化して得られた行列 $\mathbf{H} \in \mathcal{R}_+^{N\times K}$ の行ベクトル $\mathbf{h}_i$ は，K 個のコミュニティに対するノード i の属するコミュニティの分布を表す．この行ベクトルに対して KMeans 法など何らかの方法を適用し，各ノードにコミュニティラベルを割り当てる．本研究では正解ラベルとの比較のため，式 (6) により $\mathbf{h}_i$ の最大値をとるコミュニティk をノード i のラベル c_i とする．

$$c_i = \arg\max_k h_{ik} \tag{6}$$

2.2 半教師あり非負値行列分解

Yang らは式 (3) にペアワイズの制約項を加えた目的関数を最適化する半教師あり非負値行列分解を提案した [Yang 15]．まずノード i,j が同じコミュニティに属するべきという制約に対して $o_{ij} > 0$, 制約がない場合 $o_{ij} = 0$ をとる制約行列を $\mathbf{O}$ とする．o_{ij} の大きさは制約の信頼度を表す．式 (3) に制約項を加えた目的関数は式 (7) で表される．

$$\mathcal{F}_{LSE} = ||\mathbf{A} - \mathbf{W}\mathbf{H}^T||_F^2 + \lambda \mathrm{Tr}(\mathbf{H}^T\mathbf{L}\mathbf{H}) \tag{7}$$

ただし，行列 $\mathbf{L}$ は $\mathbf{O}$ のラプラシアン行列である．すなわち行列 $\mathbf{D} = [d_{ij}] \in \mathcal{R}_+^{N\times N}$ を $d_{ii} = \sum_j o_{ij}$ を満たす対角行列とすると $\mathbf{L} = \mathbf{D} - \mathbf{O}$ と表せる．また λ は非負の定数で制約が結果に及ぼす影響の大きさを調整するパラメータである．

式中の制約項は $o_{ij} > 0$ のとき $||\mathbf{h}_i - \mathbf{h}_j||$ を小さくするように作用し，ノード i,j の潜在ベクトル $\mathbf{h}_i$,$\mathbf{h}_j$ のユークリッド距離が近くなるように最適化される．

また [Yang 15] で導出されている通り, 式 (7) を解く乗法更新規則は式 (8) の通りである．

$$w_{ik} \leftarrow w_{ik}\frac{(\mathbf{A}\mathbf{H})_{ik}}{(\mathbf{W}\mathbf{H}^T\mathbf{H})_{ik}}, h_{jk} \leftarrow h_{jk}\frac{(\mathbf{A}^T\mathbf{W} + \lambda\mathbf{O}\mathbf{H})_{jk}}{(\mathbf{H}\mathbf{W}^T\mathbf{H} + \lambda\mathbf{D}\mathbf{H})_{jk}} \tag{8}$$

$\lambda = 0$ のとき式 (8) は教師なし NMF の更新規則 (4) と等価になる．

2.3 Adaptive Momentum Estimation(Adam)

勾配法で最も単純な最急勾配法では勾配に係る学習率が一定である．そのため学習の収束が安定せず，また学習の進行速度が著しく遅くなる場合がある．その問題を改善するため学習率を徐々に小さくする AdaDelta[Zeiler 12] やモーメンタムにより学習速度を改善する RMSprop[Tieleman 12] など多くの改良手法が提案されている．その中でも [Kingma 15] で提案された Adaptive Momentum Estimation(Adam) は AdaDelta と RMSprop とのハイブリッド手法で，収束までの速さや安定性が優れているとして注目されている．Adam の更新規則は式 (9) で表される．

$$\theta_{t+1} \leftarrow \theta_t - \frac{\eta}{\sqrt{\frac{v}{1-\gamma^t}} + \epsilon}\frac{m}{1-\beta^t} \tag{9}$$

式中の η は学習率，β,γ はモーメンタムの減衰パラメータ，ϵ は安定化のための微小な正数，m,v はそれぞれ式 (10)(11) で表される 1 次モーメンタムおよび 2 次モーメンタムである．

$$m \leftarrow \beta m + (1-\beta)\nabla_\theta \tag{10}$$

$$v \leftarrow \gamma v + (1-\gamma)\nabla_\theta^2 \tag{11}$$

ただし ∇_θ,∇_θ^2 はそれぞれ最適化変数 θ に対する目的関数の 1 階偏微分と 2 階偏微分である．

3. 半教師あり非負値行列分解の改善

本節では半教師あり NMF の改善手法について説明する．まず前節で説明した半教師あり NMF の制約項の問題点を指摘し，目的関数の改善を提案する．その後，NMF に勾配法を適用するための非負化手法 (i),(ii) を説明する．

(i)Projected Gradient 法で最適化する手法

(ii)$w_{ik} = |w'_{ik}|$,$h_{jk} = |h'_{jk}|$ と置換する手法

勾配法は最急勾配法 (Gradient Descent) と Adam の 2 手法を用いる．いずれの手法も TensorFlow[Ababi 15] の自動微分機能を用いて実装する．

3.1 目的関数の改善

ノード i のコミュニティは潜在ベクトル $\mathbf{h}_i$ の成分の大小関係により決まる．またそのベクトルの大きさ $||\mathbf{h}_i||$ はノード i の次数に対応し，コミュニティの決定には無関係である．しかし，式 (7) の制約項においてノード i,j に正の制約が与えられたとき，$\mathbf{h}_i$ と $\mathbf{h}_j$ のユークリッド距離が近くなるように最適化されるが，その 2 つのベクトルのユークリッド距離を小さくするということは，コミュニティの制約には関係のないノードの次数にまで制約をかけることになり，不適切であると考えられる．この目的関数において制約を増やすとコミュニティ抽出の精度が落ちることは，4.2 節の実験で確認する．

この問題を解決するために潜在行列 $\mathbf{H}$ の各行ベクトル $\mathbf{h}_i$ の和を 1 になるように正規化することを考える．正規化することでコミュニティの成分の大小関係のみが制約に依存するようになり，上述の問題を回避できることが期待される．

$\mathbf{H}$ を行方向に正規化した行列を $\hat{\mathbf{H}}$ とすると

$$\hat{\mathbf{H}} = \mathbf{V}^{-1}\mathbf{H} \tag{12}$$

とかける．ただし，行列 $\mathbf{V}$ は $[\sum_{k=1}^K h_{ik}]_{ii}$ で定義される対角行列である．

この行列に対する制約項は式 (13) のように導出される．

$$\begin{aligned} R_{LSEn} &= \frac{1}{2}\sum_{i=1}^N\sum_{j=1}^N o_{ij}\left\|\hat{\mathbf{h}}_i - \hat{\mathbf{h}}_j\right\|_2^2 \\ &= \sum_{i=1}^N \hat{\mathbf{h}}_i^T\hat{\mathbf{h}}_i d_{ii} - \sum_{i\neq j}^N \hat{\mathbf{h}}_i^T\hat{\mathbf{h}}_j o_{ij} \\ &= \mathrm{Tr}(\hat{\mathbf{H}}^T\mathbf{D}\hat{\mathbf{H}}) - \mathrm{Tr}(\hat{\mathbf{H}}^T\mathbf{O}\hat{\mathbf{H}}) \\ &= \mathrm{Tr}(\hat{\mathbf{H}}^T\mathbf{L}\hat{\mathbf{H}}) \end{aligned} \tag{13}$$

したがって，改善した目的関数は式 (14) で表される．

$$\mathcal{F}_{LSEn} = ||\mathbf{A} - \mathbf{W}\mathbf{H}^T||_F^2 + \lambda\mathrm{Tr}(\hat{\mathbf{H}}^T\mathbf{L}\hat{\mathbf{H}}) \tag{14}$$

式 (14) は形が複雑であり，乗法更新規則を導くことが困難である．そのため最適化には次に説明する勾配法ベースの手法を適用する．

3.2 勾配法による最適化

一般的に勾配法では $(-\infty, \infty)$ の区間において変数を最適化する．そのため勾配法を NMF に適用するためには，変数

の非負化処理をする必要がある．ここでは非負化手法として (i)Projected Gradient 法 (PG 法) と (ii) 変数の絶対値をとる手法を考える．非負化手法 (i) を最急勾配法に適用する方法については式 (5) において α_t を一定の学習率 α で固定すればよい．Adam に適用する方法では，式 (9) において更新後の変数が負になる場合 0 に写像する後処理を加える．すなわち，

$$w_{ik} \leftarrow \max\left(0, w_{ik} - \frac{\eta}{\sqrt{\frac{v_W}{1-\gamma^t}+\epsilon}} \frac{m_W}{1-\beta^t}\right) \tag{15}$$

$$h_{ik} \leftarrow \max\left(0, h_{ik} - \frac{\eta}{\sqrt{\frac{v_H}{1-\gamma^t}+\epsilon}} \frac{m_H}{1-\beta^t}\right) \tag{16}$$

式 (16) 中の v_W,v_H,m_W および m_H は式 (10)(11) により得る．また，行列 $\mathbf{A}$,$\mathbf{L}$ を疎行列として扱うことで $\mathcal{O}(TK(N+M))$ の計算時間で最適化可能である．ここで T は更新回数，K はコミュニティの潜在因子数，N はノード数，M はエッジ数である．行列 $\mathbf{W}$，$\mathbf{H}$ の初期値は $[0, 2(\sum_{ij} A_{ij}/(KN^2))^{1/2}]$ の一様分布より得る．

非負化手法 (ii) では，式 (7) 中の変数 $w_{ik}, h_{jk} > 0$ を $w_{ik} = |w'_{ik}|$,$h_{jk} = |h'_{jk}|$ と置換し，$[-\infty, \infty]$ の範囲で w'_{ik},h'_{jk} を勾配法により最適化する．絶対値で置換した場合，目的関数は $w'_{ik} = 0$ または $h'_{jk} = 0$ において微分不可能であり，その点においては劣微分により勾配を計算する．また手法 (i) と同様に行列 $\mathbf{A}$,$\mathbf{L}$ は疎行列として扱う．更新式は簡単なので省略する．

4. 比較実験

本節では前節の提案手法と既存手法の比較実験を説明する．まず実験の評価方法について述べ，その次に正解付きの実ネットワークを用いた実験について述べる．

4.1 評価方法

本実験で用いる半教師あり NMF 手法は，以下の通り 2 種類の勾配法と 2 種類の非負化手法の組み合わせと従来手法の 5 種類である:

1. GD_proj: 最急勾配法と (i) を組み合わせた手法
2. Adam_proj: Adam と (i) を組み合わせた手法
3. GD_abs: 最急勾配法と (ii) を組み合わせた手法
4. Adam_abs: Adam と (ii) を組み合わせた手法
5. Mult: 乗法更新規則による手法 [Yang 15]

これらの手法を正解付きの実ネットワークを用いてコミュニティ抽出の精度を測定し比較する．実験に用いたデータセットは Dolphins,Friendship,Polbooks および Polblogs の 4 種類である．データセットの詳細は表 1 の通りである．各手法により最適化された潜在行列 $\mathbf{H}$ に対し，式 (6) により各ノードに対して 1 つのコミュニティラベルを割り当てる．精度の評価には得られたコミュニティラベルと各データセットに与えられている正解ラベルとの正規化相互情報量 (NMI) を用いる．同一の正解ラベルをもつペアをランダムに選択してコミュニティ制約を付与していき，制約数の割合と精度変化の関係を検証する．[Yang 15] と同様に制約の割合の総数は式 (17) とする．

$$N_{pairs} = \sum_{k=1}^{K} N_k(N_k - 1)/2 \tag{17}$$

表 1: データセット

データセット	ノード数	エッジ数	コミュニティ数
Dolphins	62	159	2
Friendship	329	668	9
Polbooks	105	441	3
Polblogs	1490	19025	2

ここで K は正解コミュニティの数，N_k はコミュニティk のノード数を表す．実験は各条件毎に 10 回試行し，各手法の目的関数が最小の試行時の NMI で精度比較をする．

パラメータの設定について，コミュニティ数 K は表 1 に示す正解コミュニティ数に設定した．また，目的関数の式 (14) における制約の影響度のパラメータは $\lambda = 1.0$ に固定し，制約行列 $\mathbf{O}$ の非 0 要素は全て 1 とした．また最急勾配法と Adam の学習率はともに 0.01, また Adam の減衰パラメータはそれぞれ $\beta = 0.9$，$\gamma = 0.999$，安定化パラメータは $\epsilon = 1 \times 10^{-8}$ とした．勾配法の更新回数は 1000 回を上限とし，その回数までに収束しない場合は途中で打ち切る．

4.2 実ネットワークにおける実験

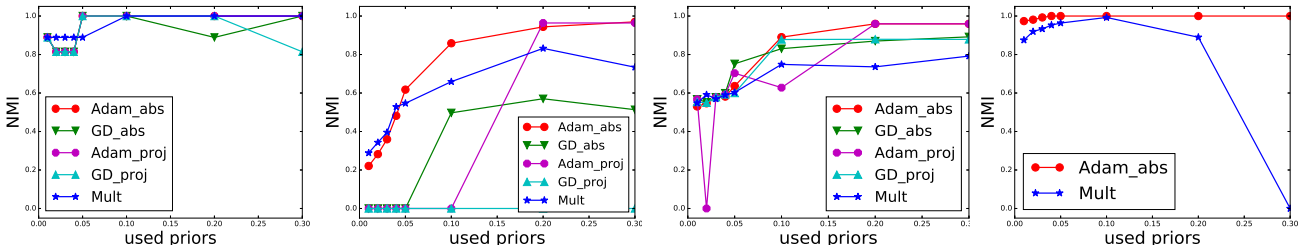
4 つの実ネットワークにおける NMI の比較結果を図 1 に示す．どのデータセットにおいても Adam と絶対値による非負化制約を組み合わせた Adam_abs が最も精度が高く安定していることがわかる．

図 1(a) の Dolphins において，Adam の 2 手法では割合 5%以降は $NMI = 1$ で安定している．一方 GD の 2 手法は制約が増えると精度が落ちている点が見られるが，これは制約が増えることによりできた関数曲面の谷に嵌り，学習が進行しなくなるからだと考えられる．図 1(b) の Friendship データセットにおいて，Adam_abs は制約が増えるにつれて既存手法の Mult よりも大きく精度が上昇している．一方 Adam_proj と GD の 2 手法は制約が少ないときに収束せず不安定であった．図 1(c) の Polbooks データセットにおいても Adam_proj は不安定で制約の割合が 1%のときは収束せず，$NMI = 0$ となった．図 1(d) の Polblogs データセットは，ノード数 1500 弱と中規模のネットワークとなっている．収束が不安定な GD_abs,GD_proj,Adam_proj は，どの制約の割合のときも収束しなかったのでグラフから除外している．Adam_abs は制約の割合の高低に関わらず高い精度を出しており，ほぼ完璧に正解コミュニティを抽出できている．一方既存手法 Mult は制約の割合が 10%までは精度が向上していたが，徐々に精度が下がり，最終的には $NMI = 0$ にまで落ちた．

4.3 考察

実験結果から各手法の比較をする．勾配法について，最急勾配法は図 1(b) の Friendship や Polblogs ネットワークにおいて，収束せず全く学習しない場合が多く見受けれた．これはコミュニティ数 K やノード数 N が大きくなり，最適化する変数の数が増え，目的関数が複雑となったことによると考えられる．これは学習率を小さくすることで回避できると考えられるが，そうすると学習の進行が遅くなり非効率的である．一方 Adam は少ない更新回数で安定した精度が出ており，より優れた勾配法であるといえる．

また図 1(b) や図 1(c) で見られるように，Adam を用いた場合でも PG 法による非負化処理をした場合，収束が不安定になった．これは勾配減算後の変数が負値となったとき，Adam のモーメンタムである式 (10) および (11) が負の値をもつ変数で計算された後に 0 へ写像されることが原因であると考え



(a) Dolphins (b) Friendship (c) Polbooks (d) Polblogs

図 1: 実ネットワークにおける NMI の比較．手法名の Adam,GD は勾配法の種類，abs,proj はそれぞれ (i) 変数の絶対値を取る手法と (ii)Projected Gradient 法を表す．Mult は乗法更新規則を用いた既存手法である．縦軸が精度 (NMI)，横軸が制約を与えたノードのペア数の割合を示す．

られる．ゆえに Adam のようにモーメンタムを扱う勾配法と Projected Gradient 法の組み合わせは相性が悪いといえる．一方絶対値による非負化手法では，Adam と組み合わせたときでも安定した精度が出ており，相性が良い組み合わせといえる．

また乗法更新規則を用いた従来手法 Mult は，図 1(b) の Friendship のようにコミュニティ数が多いネットワークや図 1(d) の Polblogs のような中規模のネットワークにおいて，制約を増やすことにより精度が落ちる現象が見受けられる．これは 3.1 節で述べたように，潜在ベクトルのユークリッド距離に対して損失が発生し，選択されたノードのペアの次数にまで制約がかかることで精度が落ちたのだと考えられる．

以上の考察より，絶対値の非負化手法 (ii) を Adam に適用した方法により，制約項を正規化した目的関数を最適化する手法 Adam_abs が精度の面において最も優れていると結論付けられる．

5. おわりに

本研究では半教師あり非負値行列分解の目的関数の改善と非負化手法を組み合わせた Adam による最適化という方法により新しい制約付きコミュニティ抽出手法を提案した．目的関数の改善では，制約項における潜在ベクトルを正規化することにより，与えたコミュニティ制約に対して精度良くコミュニティを抽出することを実現した．また，実ネットワークを用いた実験において，2 つの非負化手法 (i)Projected Gradient 法および (ii) 変数の絶対値を取る方法と最急勾配法および Adam をそれぞれ組み合わせた手法の比較を行った．その結果，絶対値による非負化手法と Adam 組み合わせた手法において，従来手法を大きく上回る精度が出ることを確認した．

本稿で提案した半教師あり NMF 手法は，勾配法によって目的関数を最適化するため，より複雑な生成モデルへの置き換えや，2 部ネットワークなどの他種類のネットワークへの拡張が容易である．それだけではなく NMF ベースの特徴抽出器やレコメンドシステムなどコミュニティ抽出以外にも応用可能であり，汎用性が高い半教師あり学習手法であると考えられる．また今後は TensorFlow で実装した提案手法を GPGPU やクラウド環境の利用によりスケーリングし，大規模データに適用可能であるかを検証していく．

参考文献

[Ababi 15] Martn Abadi, Ashish Agarwal, Paul Barham et.al. “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015. http://tensorflow.org/

[Fortunato 10] Santo Fortunato. “Community detection in graphs,” *Physics Reports*, Vol. 486, pp. 75174, 2010.

[Kingma 15] Kingma, D. P., Ba, J. L. “Adam: a Method for Stochastic Optimization” International Conference on Learning Representations, 113, 2015.

[Lee 00] D.D.Lee and H.S.Seung. “Algorithms for non-negative matrix factorization,” *Proc. Adv. Neural Inf. Process. Syst.*, pp.556-562, 2000.

[Lin 05] Lin, C.-J., “Projected gradient methods for non-negative matrix factorization” Technical Report Information and Support Services Technical Report ISSTECH-95-013, Department of Computer Science, National Taiwan University, 2005.

[Newman 04] M. E. J. Newman, “Fast algorithm for detecting community structure in networks,” *Physical Review E*, Vol.69, No. 066133, pp. 1-5, 2004.

[Psorakis 11] I. Psorakis, S. Roberts, M. Ebden, and B. Sheldon. “Overlapping community detection using Bayesian non-negative matrix factorization,” *Phisical Review E*, Vol. 83, No. 066102, 2011.

[Tieleman 12] Tieleman, T. and Hinton, G. Lecture 6.5 - “RMSProp, COURSERA: Neural Networks for Machine Learning”. Technical report, 2012.

[Yang 15] Liang Yang, Xiaochun Cao. “A Unified Semi-Supervised Community Detection Framework Using Latent Space Graph Regularization”, *IEEE Trans Cybern*, Vol. 45(11), No. 2585-98, 2015.

[Zeiler 12] Zeiler, Matthew D. “Adadelta: An adaptive learning rate method”. arXiv preprint arXiv:1212.5701, 2012.