2013 5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)

Discovery of Path-Important Nodes using Structured Semi-Nonnegative Matrix Factorization

Shawn Mankad
Decisions, Operations and Information Technologies
University of Maryland
College Park, MD 20742
Email: smankad@rhsmith.umd.edu

George Michailidis
Departments of Statistics and
Electrical Engineering & Computer Science
University of Michigan
Ann Arbor, MI 48109-1092
Email: gmichail@umich.edu

***Abstract*—Identifying critical components in networked systems is a key problem for many important applications in a diverse set of fields, including epidemiology, e-commerce and traffic systems. This paper describes the development and application of a semi-nonnegative matrix factorization for structural discovery featuring nodes that are important for transmission over social networks. The technique allows the practitioner to perform structured matrix factorization by specifying context-specific network statistics that guide the solution. The techniques are demonstrated on a network derived from Twitter data.**

I. Introduction

Through an increased understanding of how ideas are generated and disseminated, the identification of important nodes in social networks can improve many far ranging applications from targeted advertisements, to stopping the spread of epidemics. In this work, we propose a new matrix factorization-based approach for the discovery of path-important nodes within potentially weighted and directed networks. This task is different from typical community detection, which aims to extract groups of nodes that feature relatively dense within group connectivity and sparser between group connections. In contrast, the goal in this article is to discover structure featuring nodes that are important for transmission over the network. This goal is closely related to role identification, which aims to assign roles based on local connectivity patterns. Typically, role analysis methods rely on analyzing ego networks (the union of a node and its neighbors), network statistics, or graph-coloring techniques (see [1] and references therein).

The underlying methodology in this article utilizes a low rank approximation to an adjacency matrix to facilitate identification of important network substructure. The use of low rank approximations to graph related matrices follows a long line of previous works. For instance, in classical spectral layout, the coordinates of each node are given by the Singular Value Decomposition (SVD) of the Laplacian matrix. Recently, there has been extensive interest in spectral clustering (see [2] and references therein), which discovers community structure in eigenvectors of the Laplacian matrix.

Low-rank approximations satisfying different (relaxed) constraints than orthonormality are also popular. For instance, in non-negative matrix factorization (NMF), decompositions are composed of only non-negative entries. Such factorizations have been shown to be advantageous for visualization of non-negative data [3]. Note that non-negativity is typically satisfied with networks, as edges commonly correspond to flows, capacity, or binary relationships, and hence are non-negative. NMF solutions do not have simple expressions in terms of eigenvectors. They can, however, be efficiently computed by formulating the problem as one of penalized optimization, and using modern gradient-descent algorithms. Given that our proposed approach utilizes non-negativity on one of the matrix factors, we utilize a similar algorithmic approach to the NMF literature. Recently, theoretical connections between NMF and important problems in data mining have been developed [4], [5], and accordingly, NMF has been proposed for overlapping community detection on static [6], [7] and dynamic [8] networks.

The next section introduces the matrix factorization model, Section III provides estimation details, and Section IV illustrates the model on Twitter data. This article closes with a brief discussion in Section V.

II. Model

Suppose we observe a graph G, with n nodes and adjacency matrix A. We postulate the following graph Structured Semi-NMF model

$$\min_{\Lambda, V \geq 0} ||A - S\Lambda V^T||_F^2, \tag{1}$$

where $S \in \mathbb{R}^{n\times d}, \Lambda \in \mathbb{R}^{d\times K}$, and $V \in \mathbb{R}_+^{n\times K}$.

Each column of S is equal to a network statistic (degree of each node, betweenness, etc.). Hence, S is known and calculated from the graph G before performing the factorization. For instance, if the degree and betweenness are believed to be important measures for the graph G, then the S matrix would be constructed as

$$S = \begin{pmatrix} Degree(1) & Betweenness(1) \\ Degree(2) & Betweenness(2) \\ ... & ... \\ Degree(n) & Betweenness(n) \end{pmatrix}. \tag{2}$$

The S matrix utilizes information on each node that guides the factorization. Λ reveals the importance of each network statistic to the K-dimensions (columns) of V. Each node's coordinate in lower $K < n$ dimensional space is contained in V and is constrained to contain values greater than or equal to zero. The product ΛV^T provide coordinates or weights for each node that reproduce the given adjacency matrix as best as possible when left multiplied with S.

978-1-4673-3146-3/13/$31.00 ©2013IEEE

Enforcing non-negativity on a single matrix factor was first proposed in [9] with the so-called semi-NMF to improve interpretability of the resultant factorizations with data of mixed signs. We extend semi-NMF to the network setting by considering a structured approach that incorporates graph geometry into the factorization through the matrix S.

III. Algorithms

The estimation algorithm we present is an iterative one that cycles between Λ and V. The update for Λ is based on a least squares solution and derived through standard arguments. The update for V is similar to the benchmark algorithm for NMF, known as "multiplicative updating" [10]. It is obtained by utilizing the Karush-Kuhn-Tucker (KKT) optimality conditions, which provide necessary conditions for a local minimum. The derivation is omitted due to space constraints. However, the steps are similar to those found in previous works [9].

A. Updating Λ

Update Λ by setting

$$\Lambda = (S^T S)^{-1} S^T A V (V^T V)^{-1}. \tag{3}$$

Fixing V, the update for Λ is the optimal solution to $\min_\Lambda ||A - S\Lambda V||_F^2$. This can be proven by rewriting the objective function

$$\begin{aligned} \mathcal{O} &= ||A - S\Lambda V^T||_F^2 & (4)\\ &= Tr(A - S\Lambda V^T)^T (A - S\Lambda V^T), & (5) \end{aligned}$$

and computing the partial derivative

$$\frac{\partial \mathcal{O}}{\partial \Lambda} = -2S^T A V + 2(S^T S)\Lambda V^T V. \tag{6}$$

Setting the partial derivative equal to zero, fixing V and solving for Λ yields the form above.

B. Updating V

As with Λ, we can work directly with the partial derivative

$$\frac{\partial \mathcal{O}}{\partial V} = -2A^T S\Lambda + 2V\Lambda^T S^T S\Lambda \tag{7}$$

to obtain an alternating constrained least squares algorithm that boasts a quadratic convergence rate. The cost per iteration can become high, since V should be solved for subject to non-negativity using active set methods (see [11]) in

$$V\Lambda^T S^T S\Lambda = A^T S\Lambda. \tag{8}$$

Another option proposed originally for NMF by [12] is to solve for V using the usual least squares estimator

$$V = A^T S\Lambda(\Lambda^T S^T S\Lambda)^{-1}, \tag{9}$$

then set negative values in V to 0. The projection step heuristically approximates the true, non-negative solution. Notice the update for V^T after left multiplying by Λ is $\Lambda V^T = \Lambda(\Lambda^T S^T S\Lambda)^{-1}\Lambda^T S^T A$, the familiar form of predicted values from weighted least squares, where the weights are given by the S input matrix, Λ acts as the design matrix, and A is analogous to multiple responses.

In many practical situations, it may not be known which statistics, if any, are sufficient. Hence, one may be tempted to include many possible network statistics for each node. Such an approach can benefit from additional regularization in the form of a ridge (l_2) penalty to provide numerical stability. The estimator for the approximate case becomes

$$V = A^T S\Lambda(\Lambda^T S^T S\Lambda + \lambda I)^{-1}, \tag{10}$$

where λ is the regularization parameter. Other penalties on V or Λ can be investigated by utilizing results from multivariate regression, since both updates can be written as least squares problems. While easy to implement, theoretical properties are difficult to obtain due to the projection step.

A third option that is relatively simple to implement and solves for V subject to non-negativity constraints is multiplicative updating. This popular approach has been shown to converge slowly due to its linear rate [13]. However, in practice we find that after a handful of iterations, the algorithm results in meaningful factorizations.

We develop the update by introducing the Lagrangian

$$\mathcal{L}(V) = Tr(A - S\Lambda V^T)^T (A - S\Lambda V^T) - \beta V, \tag{11}$$

where β is the Lagrange multiplier enforcing non-negativity. The KKT optimality condition is obtained by setting

$$\frac{\partial \mathcal{L}}{\partial V} = -2A^T S\Lambda + 2V\Lambda^T S^T S\Lambda - \beta = 0. \tag{12}$$

Then the KKT complimentary slackness condition yields

$$(-2A^T S\Lambda + 2V\Lambda^T S^T S\Lambda)_{ij} V_{ij} = \beta_{ij} V_{ij} = 0. \tag{13}$$

Separate the positive and negative parts of a matrix X as $X_{ij}^+ = (|X_{ij}| + X_{ij})/2$ and $X_{ij}^- = (|X_{ij}| - X_{ij})/2$. Then the following update satisfies fixed point Equation 13

$$V_{ij} = V_{ij} \sqrt{\frac{(A^T S\Lambda)_{ij}^+ + (V\Lambda^T S^T S\Lambda)_{ij}^-}{(A^T S\Lambda)_{ij}^- + (V\Lambda^T S^T S\Lambda)_{ij}^+}}. \tag{14}$$

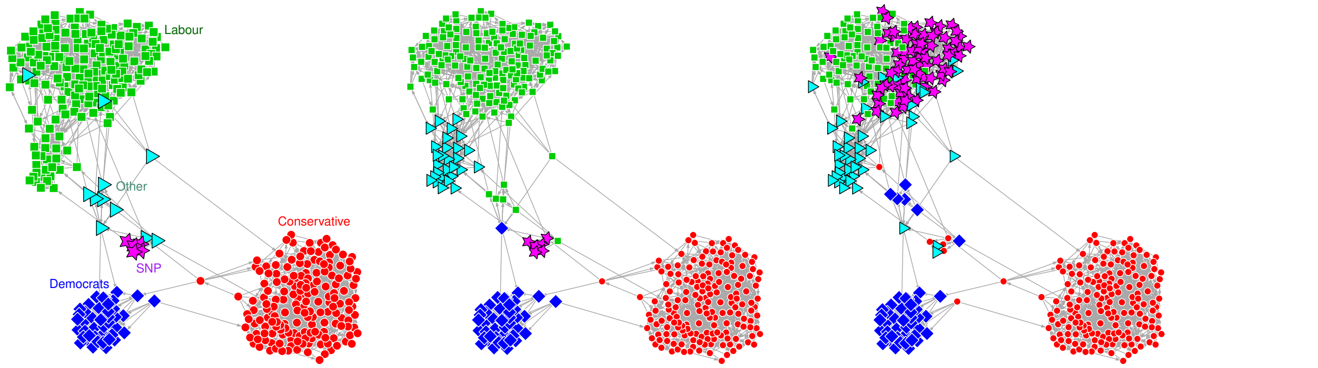
A theoretical property of Equation 14 is that it is non-increasing with respect to the objective function. Details proving this property can be found in [9].

IV. UK Parliament Twitter Network

We analyze a Twitter network of 419 Members of Parliament (MPs) in the United Kingdom. The raw data, collected and processed by the authors in [14], consists of approximately 540,000 tweets, 3,000 user lists, and 27,000 follower links within the set of 419 MPs from late 2012. The different aspects of Twitter data are integrated using the techniques in [14] to produce a single unified graph, shown in Fig. 1(a), that we analyze.

The network features strong community structure, corresponding to different political affiliations of the MPs. As such, community detection methods, such as spectral clustering [2], classical and Semi-NMF [6], [7]), recover the true structure fairly well with the exception of the Labour party, as seen in Fig. 1(b,c). Both approaches tend to split the Labour party into sub-groups corresponding to known intraparty divisions[1].

[1] See http://blogs.telegraph.co.uk/news/danhodges/100134777/the-division-of-labour-todays-party-hasnt-got-a-clue-where-its-going/



(a) Observed Network (b) Spectral Clustering (c) Semi-NMF

Fig. 1. Subfigure (a) shows the network of UK Members of Parliament, with color and vertex shapes denoting party affiliation. Subfigures (b) and (c) display typical results of applying spectral clustering and Semi-NMF to the UK-MP Twitter network. Both algorithms recover well the true political party structure.

These community detection methods can help guide other approaches for finding path important MPs. For instance, a practitioner could first discover communities, then search for interesting network statistic profiles. We present in Fig. 2(b) the subgraph of highest degree MPs within each party. Fig. 2(c) shows the subgraph of MPs within each party with the largest magnitude in the first principal component of clustering coefficient, betweenness, and closeness. These three network statistics measure different aspects of local geometry and shortest path properties. In both subfigures, we see isolated nodes and disconnected communities. We also note that both of these direct approaches utilize the true community membership to illustrate the most opportunistic set of results.

The Structured Semi-NMF does not utilize party affiliations, yet results in more interpretable subgraphs. Shown in Fig. 2(a), there are no isolated nodes, links within each community are more dense, and connections between communities exist. Statistics measuring connectedness in Table I indicate that the proposed approach performs best. It has the least fragmentation, the largest connected subgraph, and highest clustering coefficient.

The nodes comprising the subgraph are chosen by keeping the group with largest mean after applying K-Means ($K = 2$) on V, with $rank(V) = 2$ and $S =$ [Clustering Coef.,Betweenness, Closeness]. Results are similar for other V ranks. There are a number of guidelines we consider when choosing parameters for the proposed factorization. First, due to the weight matrix S, the product $S\Lambda V^T$ can be at most $rank(S)$, which is equal to three and not large in general. Accordingly, as shown in Fig. 3, there is little improvement in explained variance for additional dimensions in V after three. Second, the variables in S are chosen to be path-related and not direct functions of eigenvectors. As a consequence, the factorization features higher error compared to classical NMF. The benefit of the Structured Semi-NMF is that network structure featuring path important MPs are highlighted instead of party affiliations.

V. Conclusion

The graph Structured Semi-NMF method appears to outperform competing approaches that either utilize network statistics only or exclude S in matrix factorization for identification of transmission substructure. The extra flexibility of semi-NMF along with carefully chosen network statistics steer the factorization towards highly interpretable solutions. Moreover, the factorization can be extended by constructing S from potentially available meta data, thus providing a way to combine network data with additional vertex features.

A weakness of the approach is illustrated in situations when it is unclear which network statistics, if any, are the right ones. The usefulness of graph Structured Semi-NMF depends upon the practitioner choosing appropriate, context-specific statistics. When the S matrix is high dimensional, the additional regularization discussed briefly above for the

TABLE I. Network statistics of the different subgraphs shown in Fig. 2. Number Components refers to the number of strongly connected components, LSCC refers to the proportion of nodes in the largest strongly connected component.

	Number Components	LSCC	Clustering Coefficient
Structured Semi-NMF	3	0.59	0.66
Highest Degree	4	0.45	0.48
Largest Path	9	0.45	0.55
Random	28	0.24	0.09

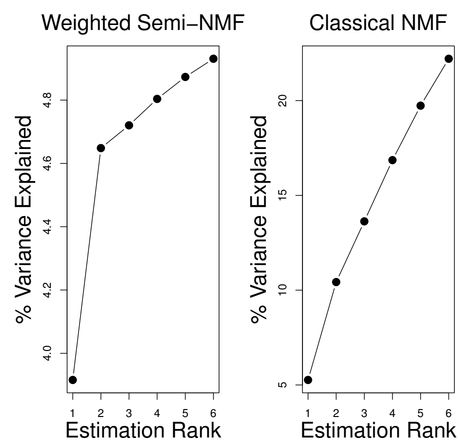
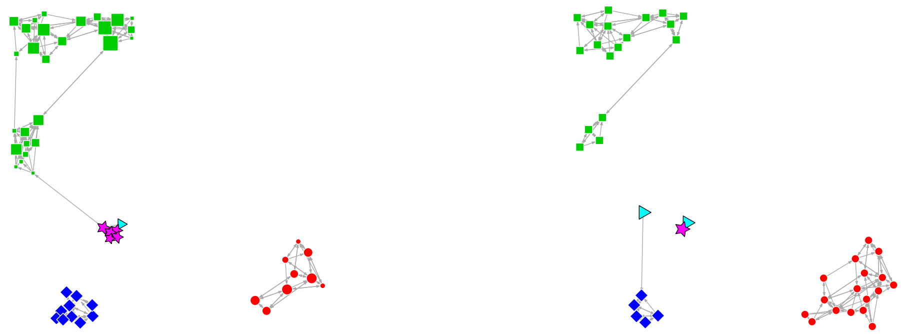
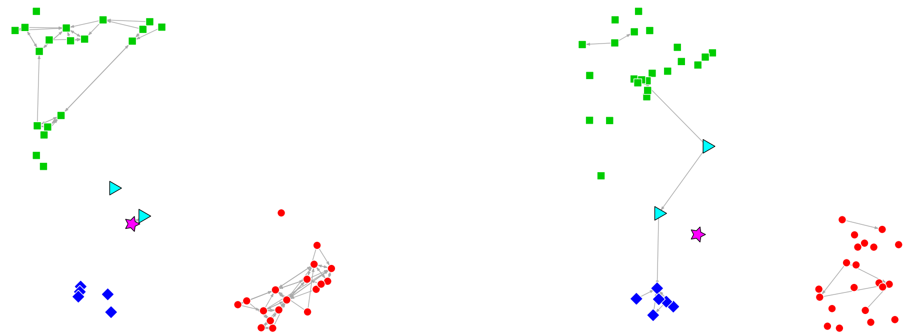


Fig. 3. Percentage of variance explained $(1 - ||A - \hat{A}||_F^2/||A||)F^2)$ of the graph Structured Semi-NMF and that of classical NMF for the UK Parliament Twitter Network.



(a) Structured Semi-NMF subgraph (unsupervised)   (b) Highest degree nodes within each party (supervised)



(c) Largest path statistic nodes within each party (supervised)   (d) Random party subgraphs (supervised)

Fig. 2. Subfigure (a) shows the subgraph after applying K-Means to V from Structured Semi-NMF. Node sizes proportional to their norm in V, which is of rank 2. Subfigures (b),(c) and (d) shows the subgraph of nodes according to different node selection schemes that condition on the true party affiliations of the MPs.

alternating least squares algorithms should be further studied.

References

[1] U. Brandes, J. Lerner, and U. Nagel, "Network ensemble clustering using latent roles," *Adv. Data Anal. Classif.*, vol. 5, no. 2, pp. 81–94, Jul. 2011. [Online]. Available: http://dx.doi.org/10.1007/s11634-010-0074-3

[2] K. Rohe and B. Yu, "Co-clustering for Directed Graphs; the Stochastic Co-Blockmodel and a Spectral Algorithm," *ArXiv e-prints*, Apr. 2012.

[3] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 10 1999.

[4] C. Ding, X. He, and H. D. Simon, "On the equivalence of nonnegative matrix factorization and spectral clustering," in *Proc. SIAM Data Mining Conf*, 2005, pp. 606–610.

[5] C. Ding, T. Li, and W. Peng, "On the equivalence between non-negative matrix factorization and probabilistic latent semantic indexing," *Comput. Stat. Data Anal.*, vol. 52, no. 8, pp. 3913–3927, Apr. 2008. [Online]. Available: http://users.cis.fiu.edu/~taoli/pub/NMFpLSIequiv.pdf

[6] I. Psorakis, S. Roberts, M. Ebden, and B. Sheldon, "Overlapping community detection using bayesian non-negative matrix factorization," *Phys. Rev. E*, vol. 83, p. 066114, Jun 2011. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevE.83.066114

[7] F. Wang, T. Li, X. Wang, S. Zhu, and C. Ding, "Community discovery using nonnegative matrix factorization," *Data Mining and Knowledge Discovery*, vol. 22, no. 3, pp. 493–521, 2011. [Online]. Available: http://dx.doi.org/10.1007/s10618-010-0181-y

[8] Y.-R. Lin, Y. Chi, S. Zhu, H. Sundaram, and B. L. Tseng, "Facetnet: a framework for analyzing communities and their evolutions in dynamic networks," in *Proceedings of the 17th international conference on World Wide Web*, ser. WWW '08. New York, NY, USA: ACM, 2008, pp. 685–694. [Online]. Available: http://doi.acm.org/10.1145/1367497.1367590

[9] C. Ding, T. Li, and M. Jordan, "Convex and semi-nonnegative matrix factorizations," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 1, pp. 45–55, 2010.

[10] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *Advances in neural information processing systems*, pp. 556–562, 2001. [Online]. Available: hebb.mit.edu/people/seung/papers/nmfconverge.pdf

[11] H. Kim and H. Park, "Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method," *SIAM Journal on Matrix Analysis and Appcliations*, vol. 30, no. 2, pp. 713–730, 2008.

[12] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization," in *Computational Statistics and Data Analysis*, 2006, pp. 155–173.

[13] M. Chu, F. Diele, R. Plemmons, and S. Ragni, "Optimality, computation, and interpretation of nonnegative matrix factorization," *SIAM Journal of Matrix Analysis*, pp. 4–8030, 2004. [Online]. Available: http://users.wfu.edu/plemmons/papers/chu_ple.pdf

[14] D. Greene and P. Cunningham, "Producing a Unified Graph Representation from Multiple Social Network Views," *ArXiv e-prints*, Jan. 2013.