Normalized Dominant Set on the Vertex- and Edge-Weighted Graph

Xianglong Liu
Beihang University
xlliu@nlsde.buaa.edu.cn

Junfeng He
Columbia University
jh2700@columbia.edu

January 13, 2013

1 Graph-based Definition

Motivated by recent research on modelling correlations based on graph [1, 2], we first introduce a vertex-weighted and undirected edge-weighted graph: $G = (V, E, A, \pi)$, where $V = \{1, \ldots, L\}$ is the vertex set corresponding to the L types of entities with weights $\pi = [\pi_1, \ldots, \pi_L]^T$, and $E \subseteq V \times V$ is the edge set with weights $A = (a_{ij})$. Each $a_{ij} : (i, j) \in E \rightarrow \mathbb{R}_+$, a positive weight corresponding to the edge between vertex i and j.

Based on the graph representation, a important problem is a dominant set discovery problem [2], but in the vertex- and edge-weighted graph G. In the literature there are very few work that uncover the most dominant set on a vertex- and edge-weighted graph. In this section, we will introduce the concept of normalized dominant set on such graph, and prove that the local optima of a quadratic programming defined in (4) corresponds to the normalized dominant set.

2 Normalized Dominant Set

The normalized dominant set should have high vertex and edge weights inside, *i.e.*, high internal homogeneity and high inhomogeneity between vertices inside and those outside. Following prior successful research on dominant set [1, 2], we start with an idea that both vertex and edge weights will induce a discriminative assignment of weights on the vertices, with respect to the homogeneity among the whole vertex set.

2.1 relative internal homogeneity

Let $S \subseteq V$ be a nonempty subset of vertices and $j \in S$. We attempt to characterize the induced vertex weights by measuring the connection from any vertex $j \in S$ to $i \notin S$. It should take both the internal homogeneity in S and the vertex weights of j and i into account. We first define a function $f(S, j|i)$ for the relative internal homogeneity between j and other vertices in S with respect to i. It should satisfy three basic properties: (1) **Non-negativity** f should be non-negative for S, j, and i; (2) **Symmetry** f should be symmetric with respect to all vertices in S; and (3) **Monotonicity** f should be monotonic in each of its arguments.

Positivity and symmetry are simple consequences of the internal homogeneity definition. Monotonicity is a reasonable constraint from the following aspects: f should be monotonically increasing with respect to S and j, because the increase of vertex weighs and edge weights in S will lead to higher internal homogeneity between j and S; f should be monotonically decreasing with respect to vertex i, due to the intuition that a large weight on i indicates relatively small internal homogeneity between j and S.

With these properties, there are many choices for f, and in this paper we consider the following form:

$$f(S, j|i) = \frac{\pi_i^{-1}}{\sum_{k \in S} \pi_k^{-1}} \sum_{k \in S} a_{jk}. \tag{1}$$

Note that when π are all identical, f will degenerate to the average weight between j and S.

2.2 relative external connection

Now we can define the connection between vertex $j \in S$ and $i \notin S$ motivated by the transition rate in Markov Jump:

$$\phi_S(j, i) = \frac{\pi_j}{\pi_i}(a_{ji} - f(S, j|i)). \tag{2}$$

$\phi_S(j, i)$ measures the relative external connection from vertex $j \in S$ to $i \notin S$, considering not only the vertex- and edge weights between vertex j and i, but also the internal homogeneity of j in S.

The connection strength is determined by both $(a_{ji} - f(S, j|i))$, the homogeneity between j and i eliminating the internal homogeneity of i in S, and the vertex weight ratio $\frac{\pi_j}{\pi_i}$. If π_j is small compared to π_i, vertex j will contribute less to vertex i. See the illustrative example in Figure 1 (a). Note $\phi_S(j, i)$ can be either positive or negative.

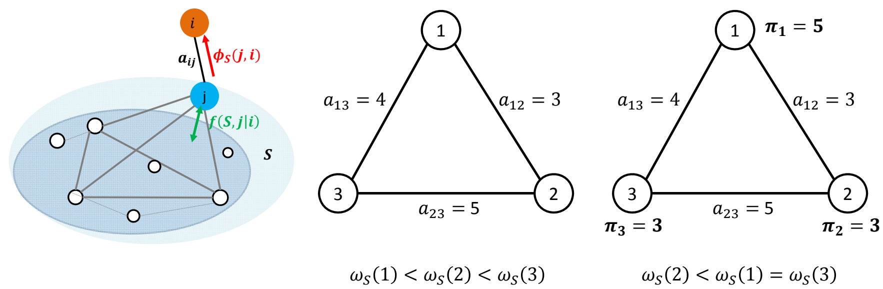
2.2.1 induced vertex weights

Now we formalize the induced vertex weights in a recursive way:

Definition 1 *Let $S \subseteq V$ be a nonempty vertex subset and $i \in S$. The induced vertex weight of i with regard to S is*

$$w_S(i) = \begin{cases} \frac{1}{\pi_i^2}, & \textit{if } |S| = 1 \\ \sum_{j \in S \setminus \{i\}} \phi_{S \setminus \{i\}}(j, i) w_{S \setminus \{i\}}(j), & \textit{otherwise.} \end{cases} \tag{3}$$

The total weight of S is defined to be: $W(S) = \sum_{i \in S} w_S(i)$. The induced vertex weight $w_S(i)$ serves as a measure of the relative overall connections between vertex i and the remaining vertices in S. Therefore it can be naturally regarded as a rank score for each vertex, normalized by the vertex weights in the computation.



(a) (b) (c)

Figure 1: (a) An illustration of relative connection defined in (2). The induced vertex weight of vertex i with regard to $S=\{1,2,3\}$ for two types of graph: (b) Edge-weighted graph [2]: $w_S(1)=10<w_S(2)=16<w_S(3)=18$; (c) Vertex-weighted Edge-weighted graph: $w_S(2)=\frac{8}{9}<w_S(1)=\frac{4}{3}=w_S(3)=\frac{4}{3}$.

An example comparing our induced vertex weight and that of the dominant set [2] is shown in Figure 1. As we can see, our induced vertex weights, aggregating both vertex and edge weights, alter the final rank of vertices, and assign high scores to vertices with large vertex weights.

2.3 definition

We introduce the definition of the normalized dominant set in a vertex- and edge-weighted graph based on induced vertex weights:

Definition 2 *A nonempty subset of vertices $S\subseteq V$ such that $W(T)>0$ for any nonempty $T\subseteq S$, is said to be normalized dominant if: (1) $w_S(i)>0$, for all $i\in S$; (2) $w_{S\cup\{i\}}(i)\le 0$, for all $i\notin S$.*

The first condition of the above definition forces the strong connections among vertices in the normalized dominant set, while the second regards external inhomogeneity.

3 Solution

Motivated by previous research on subset selection [1,2], it can be solved by a quadratic programming with respect to $\mathbf{x}$. Each element of $\mathbf{x}$ expresses the importance of association with the desired subset. Therefore, if define the support of $\mathbf{x}$ as $\sigma(\mathbf{x})=\{i\in V: x_i\neq 0\}$, then the desired subset corresponds to the elements in the support with largest values. This turns to a quadratic programming with continuous constraints on $\mathbf{x}$:

$$\begin{array}{ll}\max & \frac{1}{2}\mathbf{x}^T\hat{A}\mathbf{x}\\ \text{s.t.} & \mathbf{x}\in\Delta\end{array}\tag{4}$$

where

$$\Delta=\{\mathbf{x}\in\mathbb{R}^L:\mathbf{x}\ge 0 \text{ and } \mathbf{1}^T\mathbf{x}=1\}\tag{5}$$

with $\hat{A}$ defined as

$$\hat{A} = \Pi A \Pi, \tag{6}$$

where $\Pi = \text{diag}(\pi)$.

A straightforward and powerful way to find (local) solutions of a quadratic programming problem is the so-called replicator dynamics [1, 2], arising in evolutionary game theory. Given an initialization of $\mathbf{x}(0)$, the iteration can be performed efficiently in the following form:

$$x_i(t+1) = x_i(t)\frac{(\hat{A}\mathbf{x}(t))_i}{\mathbf{x}(t)^T\hat{A}\mathbf{x}(t)},\ i = 1, \ldots, L. \tag{7}$$

Under these dynamics, the simplex Δ is invariant. Moreover, it has been proven that with symmetric $\hat{A}$ whose entries are nonnegative, the objective function will strictly increase, and its asymptotically stable points correspond to strict local solutions [1].

We establish the intrinsic connections between the normalized dominant set and the local optima of quadratic programming in (4) by the following theorem.

Theorem 1 *If $\mathbf{x}^*$ is a strict local solution of program (4) with $\hat{A} = \Pi A \Pi$, where $\Pi = diag(\pi)$, then its support $\sigma = \sigma(\mathbf{x})$ is the normalized dominant set of graph $G = (V, E, A, \pi)$, provided that $w_{\sigma\cup\{i\}}(i) \neq 0$ for all $i \notin \sigma$.*

See the supplementary materials for proofs.

3.1 Proof of Theorem 1

To prove the theorem in the paper, we first prove the following proposition and lemmas.

The definition of the normalized dominant set can be characterized and represented in terms of determinants. For $S \subseteq V$, we denote by A_S the submatrix of A formed by the rows and the columns indexed by the elements of S. Then we define the matrix B_S:

$$B_S = \begin{pmatrix} 0 & (\pi^{-1})^T \\ \pi^{-1} & A_S \end{pmatrix}, \tag{8}$$

and the matrix ${}^jB_S$:

$${}^jB_S = \begin{pmatrix} 0 & & & (\pi^{-1})^T & & & \\ \pi^{-1} & A_S^1 & \ldots & A_S^{j-1} & \mathbf{0} & A_S^{j+1} & \ldots & A_S^m \end{pmatrix}, \tag{9}$$

where $S = \{i_1, \ldots, i_m\}$ with $i_1 < \ldots < i_m$ and A_S^i is the i-th column of A_S.

Proposition 1 *Let $S = \{i_1, \ldots, i_m\} \subseteq V$ be a nonempty subset of vertices and assume $i_1 < \ldots < i_m$ without loss of generality. Then for any $i_h \in S$,*

$$w_S(i_h) = (-1)^m \det({}^hB_S), \tag{10}$$

and,

$$W_S = (-1)^m \det(B_S). \tag{11}$$

Proof First we prove (10) holds for $m \geq 1$.

1. For $m = 1$,

$$ {}^1B_{\{i\}} = \begin{pmatrix} 0 & \pi_i^{-1} \\ \pi_i^{-1} & 0 \end{pmatrix}, $$

It is easy to verify that $w_{\{i\}}(i) = \frac{1}{\pi_i^2} = -\det({}^1B_{\{i\}})$.

2. For $m > 1$ and $S = \{i_1, \ldots, i_m\}$ with $i_1 < \ldots < i_m$,

$$
\begin{aligned}
\det({}^hB_S) &= \begin{vmatrix} 0 & \pi_{i_1}^{-1} & \cdots & \pi_{i_h}^{-1} & \cdots & \pi_{i_m}^{-1} \\ \pi_{i_1}^{-1} & a_{i_1 i_1} & \cdots & 0 & \cdots & a_{i_1 i_m} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ \pi_{i_h}^{-1} & a_{i_h i_1} & \cdots & 0 & \cdots & a_{i_h i_m} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ \pi_{i_m}^{-1} & a_{i_m i_1} & \cdots & 0 & \cdots & a_{i_m i_m} \end{vmatrix} \\
&= \frac{(-1)^{h+2}}{\pi_{i_h}} \begin{vmatrix} \pi_{i_1}^{-1} & a_{i_1 i_1} & \cdots & a_{i_1 i_m} \\ \vdots & \vdots & \cdots & \vdots \\ \pi_{i_h}^{-1} & a_{i_h i_1} & \cdots & a_{i_h i_m} \\ \vdots & \vdots & \cdots & \vdots \\ \pi_{i_m}^{-1} & a_{i_m i_1} & \cdots & a_{i_m i_m} \end{vmatrix} \\
&= -\frac{1}{\pi_{i_h}}[\sum_{i_j \in S \backslash \{i_h\}} \pi_{i_j} a_{i_h i_j} \det({}^jB_{S \backslash \{i_h\}}) \\
&\quad + \frac{1}{\pi_{i_h}} \det(A_{S \backslash \{i_h\}})]
\end{aligned}
$$

From the fact that

$$ \begin{vmatrix} (\pi^{-1})^T\mathbf{1} & (A_S\mathbf{1})^T \\ \pi^{-1} & A_S \end{vmatrix} = 0, $$

we can obtain:

$$ \sum_{i_j \in S} \frac{1}{\pi_{i_j}} \det(A_S) + \sum_{i_j \in S} (\sum_{i_k \in S} a_{i_j i_k}) \pi_{i_j} \det({}^jB_S) = 0, $$

namely

$$ \det(A_S) = -\pi_{i_h} \sum_{i_j \in S} \pi_{i_j} f(S, i_j | i_h) \det({}^jB_S). $$

Therefore, we can rewrite $\det({}^hB_S)$ as

$$
\begin{aligned}
&- \sum_{i_j \in S \backslash \{i_h\}} \frac{\pi_{i_j}}{\pi_{i_h}} (a_{i_h i_j} - f(S \backslash \{i_h\}, i_j | i_h)) \det({}^jB_{S \backslash \{i_h\}}) \\
&= - \sum_{i_j \in S \backslash \{i_h\}} \phi_{S \backslash \{i_h\}}(i_j, i_h) \det({}^jB_{S \backslash \{i_h\}}).
\end{aligned}
$$

According to the recursive definition of $w_S(i_h)$, we can conclude that (10) holds for $m \geq 1$.

For (11), since $\det(B_S) = \sum_{i_h \in S} \det({}^h B_S)$, then $W(S) = (-1)^m \det(B_S)$. $\square$

When using the following fact

$$\begin{vmatrix} \pi_h^{-1} & (A_S^h)^T \\ \pi^{-1} & A_S \end{vmatrix} = 0,$$

we give an alternative way to compute $w_S(i)$:

$$w_S(i) = \sum_{j \in S \setminus \{i\}} \frac{1}{\pi_i^2} (\pi_i \pi_j a_{ij} - \pi_h \pi_j a_{hj}) w_{S \setminus \{i\}}(j) \tag{12}$$

with h as an arbitrary element of $S \setminus \{i\}$, and $|S| > 1$.

Lemma 1 *With $\hat{A} = \Pi A \Pi$, where $\Pi = diag(\pi)$, the KKT equality conditions in (7) in the paper hold, if and only if*

$$B_\sigma [-\lambda, \pi_{i_1} x_{i_1}, \ldots, \pi_{i_h} x_{i_h}]^T = [1, 0, \ldots, 0]^T, \tag{13}$$

where λ is a real constant number, $\sigma = \sigma(\mathbf{x}) = \{i_1, \ldots, i_h\}$ with $i_1 < \ldots < i_h$, and the matrix B_σ:

$$B_\sigma = \begin{pmatrix} 0 & \pi_{i_1}^{-1} & \ldots & \pi_{i_h}^{-1} \\ \pi_{i_1}^{-1} & a_{i_1 i_1} & \ldots & a_{i_h i_h} \\ \vdots & \vdots & \ldots & \vdots \\ \pi_{i_h}^{-1} & a_{i_h i_1} & \ldots & a_{i_h i_h} \end{pmatrix}.$$

Proof $\mathbf{x} \in \Delta$ satisfies the Karush-Kuhn-Tucker (KKT) conditions for problem (7), if there exist $n+1$ real constants (Lagrange multipliers) $\mu_1, \ldots, \mu_L$ and λ, with $\mu_i \geq 0$ for all $i = 1 \ldots L$, such that for all $i = 1 \ldots L$:

$$\begin{aligned} (\hat{A}\mathbf{x})_i - \lambda + \mu_i &= 0; \\ x_i \mu_i &= 0. \end{aligned}$$

Because of the nonnegativity of both x_i and μ_i, it can be restated as follows:

$$(\hat{A}\mathbf{x})_i \begin{cases} = \lambda, & \text{if } i \in \sigma(x); \\ \leq \lambda, & \text{otherwise} \end{cases}$$

with some real constant $\lambda = x^T \hat{A} x$.

For $\sigma = \sigma(\mathbf{x}) = \{i_1, \ldots, i_m\}$ with $i_1 < \ldots < i_m$.

$$B_\sigma = \begin{pmatrix} 0 & \pi_{i_1}^{-1} & \ldots & \pi_{i_m}^{-1} \\ \pi_{i_1}^{-1} & & & \\ \vdots & & A_\sigma & \\ \pi_{i_m}^{-1} & & & \end{pmatrix}.$$

Then $B_\sigma[-\lambda, \pi_{i_1}x_{i_1}, \ldots, \pi_{i_m}x_{i_m}]^T = [1, 0, \ldots, 0]^T$ is equivalent to:

$$\begin{cases} \sum_{h=1}^m x_{i_h} & = 1; \\ A_\sigma[\pi_{i_1}x_{i_1}, \ldots, \pi_{i_m}x_{i_m}]^T & = \lambda[\pi_{i_1}^{-1}, \ldots, \pi_{i_m}^{-1}] \end{cases}$$

namely, using $\Pi = \mathrm{diag}(\pi)$

$$\begin{cases} \mathbf{1}^T\mathbf{x} & = 1; \\ (\Pi A \Pi \mathbf{x})_i & = \lambda,\ i \in \sigma(x) \end{cases}$$

By setting $\hat{A} = \Pi A \Pi$, we prove Lemma 1. $\square$

Lemma 2 *Let $\sigma = \sigma(\mathbf{x})$ be the support of a vector $\mathbf{x} \in \Delta$. Then, $\mathbf{x}$ satisfies the KKT equality conditions in (14) in the paper if and only if*

$$x_i = \begin{cases} \frac{w_\sigma(i)}{W(\sigma)}, & \textit{if } i \in \sigma; \\ 0, & \textit{otherwise.} \end{cases} \tag{14}$$

Moreover,

$$\frac{w_{\sigma\cup\{j\}}(j)}{W(\sigma)} = \frac{1}{\pi_j^2}[(\hat{A}\mathbf{x})_j - (\hat{A}\mathbf{x})_i] = -\frac{1}{\pi_j^2}\mu_j \tag{15}$$

for all $i \in \sigma$ and $j \notin \sigma$, where the μ_j are the (nonnegative) Lagrange multipliers of program (7).

Proof For (13) which is equivalent to the KKT conditions $\hat{A}\mathbf{x})_i = \lambda$, if $i \in \sigma(x)$, it can be treated as a linear equation problem with unknowns λ and $x_i, i \in \sigma$. Since $\det(B_\sigma) \neq 0$, the problem has an unique solution whose support denoted by $\sigma = \{i_1, \ldots, i_m\}$ without loss of generality. Using Cramer's rule, we can get

$$\pi_{i_h}x_{i_h} = \frac{\pi_{i_h}\det({}^hB_\sigma)}{\det(B_\sigma)},$$

Then according to Lemma 1, we have

$$x_{i_h} = \frac{(-1)^m w_\sigma(i_h)}{(-1)^m W(\sigma)} = \frac{w_\sigma(i_h)}{W(\sigma)}.$$

for any $1 \leq h \leq m$. Therefore, $x = x^\sigma$.

Using Equation (12) in the paper, we obtain:

$$\begin{aligned} \frac{w_{\sigma\cup\{j\}}(j)}{W(\sigma)} &= \frac{\sum_{i_h\in\sigma}\frac{1}{\pi_j^2}(\pi_j\pi_{i_h}a_{ji_h} - \pi_{i_k}\pi_{i_h}a_{i_ki_h})w_\sigma(i_h)}{W(\sigma)} \\ &= \frac{1}{\pi_j^2}\sum_{i_h\in\sigma}(\pi_j\pi_{i_h}a_{ji_h} - \pi_{i_k}\pi_{i_h}a_{i_ki_h})x_{i_h}^\sigma \\ &= \frac{1}{\pi_j^2}[(\hat{A}x)_j - (\hat{A}x)_{i_k}]. \end{aligned}$$

We have the fact $(\hat{A}x)_j - (\hat{A}x)_i = -\mu_j$ for all $i \in \sigma$ and $j \notin \sigma$, and $\pi_j > 0$ for all j. Then we can conclude that

$$\frac{w_{\sigma \cup \{j\}}(j)}{W(\sigma)} = \frac{1}{\pi_j^2}[(\hat{A}x)_j - (\hat{A}x)_i] = -\frac{1}{\pi_j^2}\mu_j \leq 0.$$

for all $i \in \sigma$ and $j \notin \sigma$, where the μ_j are the (nonnegative) Lagrange multipliers of quadratic programming problem in the paper. □

Using Lemma 1 and 2, the proof can be completed following that of [2]. □

References

[1] H. Liu and S. Yan. Robust graph mode seeking by graph shift. In *ICML*, pages 671–678, 2010. 1, 3, 4

[2] M. Pavan and M. Pelillo. Dominant sets and pairwise clustering. *IEEE Trans. Pattern Anal. Mach. Intell.*, pages 167–172, 2007. 1, 3, 4, 8