Meaning of the norm Eq. 7

Since \( p_1(x, y) \) is by definition a stochastic matrix with stationary distribution \( \mu \), it admits a spectral decomposition similar to Eq. 6:

\[
p_1(x, y) = \sum_{k=0}^{n-1} \lambda_k' \varphi'_k(x) \varphi'_k(y) \mu(y).
\]

for some appropriate eigenvectors \( \varphi'_k(x) \) and eigenvalues \( \lambda'_k \). Inserting this spectral decomposition into Eq. 7, and using the orthogonality of the eigenvectors \( \varphi_k(x) \), i.e. \( \sum_{x \in S} \varphi'_k(x) \varphi'_q(x) \mu(x) = \delta_{k,q} \), we deduce that

\[
\|p_1\|_\mu^2 = \sum_{x,y} \mu(x) \mu(y) \sum_{k,q=1}^n \lambda_k' \lambda_q' \varphi'_k(x) \varphi'_k(y) \varphi'_q(x) \varphi'_q(y) = \sum_{k=1}^n \lambda_k^2
\]

Derivation of equation Eq. 12

Recall the definition

\[
\bar{p}(x, y) = \sum_{i,j=1}^N \mathbf{1}_{S_i}(x) \bar{p}(S_i, S_j) \mu_j(y),
\]

and the notation

\[
\mu_j(y) = \frac{\mu(y) \mathbf{1}_{S_j}(y)}{\bar{\mu}(S_j)}, \quad \bar{\mu}(S_j) = \sum_{y \in S_j} \mu(y).
\]

Define the function

\[
F = \sum_{x,y \in S} \frac{\mu(x)}{\mu(y)} |\bar{p}(x, y) - p_t(x, y)|^2,
\]
We wish to minimize $F$ over all $\hat{p}(S_i, S_j)$. We have

$$\frac{\partial F}{\partial \hat{p}(S_k, S_l)} = 2 \sum_{x,y \in S} \frac{\mu(x)}{\mu(y)} (\hat{p}(x,y) - p_t(x,y)) \frac{\partial \hat{p}(x,y)}{\partial \hat{p}(S_k, S_l)}$$

$$= 2 \sum_{x,y \in S} \frac{\mu(x)}{\mu(y)} \left( \sum_{i,j=1}^{N} 1_{S_i}(x) \hat{p}(S_i, S_j) \mu_j(y) - p_t(x,y) \right) 1_{S_k}(x) \mu_l(y)$$

$$= 2 \sum_{x \in S_k, y \in S_l} \frac{\mu(x)}{\mu(S_i)} \left( \hat{p}(S_k, S_l) \frac{\mu(y)}{\mu(S_l)} - p_t(x,y) \right) \mu_l(y)$$

(40)

From the optimality condition $\partial F / \partial \hat{p}(S_k, S_l) = 0$ we can obtain

$$\hat{p}(S_k, S_l) = \frac{1}{\hat{\mu}(S_k)} \sum_{x \in S_k, y \in S_l} \mu(x)p_t(x,y) = \sum_{x \in S_k, y \in S_l} \mu_k(x)p_t(x,y).$$

(41)

This is Eq. 12.

**Some properties of Eqs. 12 and 15**

The matrix $\hat{p}^*(S_i, S_j)$ defined in Eq. 12 is a stochastic matrix on $S$ since $\hat{p}^*(S_i, S_j) \geq 0$ for all $i,j = 1, \ldots, N$ and

$$\sum_{j=1}^{N} \hat{p}^*(S_i, S_j) = \sum_{j=1}^{N} \sum_{\substack{x \in S_i \ y \in S_j}} \mu_i(x)p_t(x,y)$$

$$= \sum_{x \in S_i} \mu_i(x) \sum_{y \in S} p_t(x,y) = 1.$$

(42)

In addition $\hat{\mu}(S_i)$ is an equilibrium distribution for the Markov chain on $S$ with transition matrix $p^*$ since $\hat{\mu}(S_i) \geq 0$ and

$$\sum_{i=1}^{N} \hat{\mu}(S_i) = \sum_{i=1}^{N} \sum_{x \in S_i} \mu(x) = \sum_{x \in S} \mu(x) = 1.$$

(43)
and
\[
\sum_{i=1}^{N} \hat{\mu}(S_i) \hat{p}^*(S_i, S_j) = \sum_{i=1}^{N} \sum_{x \in S_i} \mu(x) \sum_{y \in S_i} \mu_i(y) p_t(y, z) \\
= \sum_{i=1}^{N} \sum_{y \in S_i} \mu(y) p_t(y, z) \\
= \sum_{y \in S_j} \mu(y) p_t(y) = \hat{\mu}(S_j).
\]

(44)

Using the spectral decomposition (Eq. 6), note that Eq. 12 can also be expressed as
\[
\hat{p}^*(S_i, S_j) = \sum_{k=0}^{n-1} \lambda_k \tilde{\varphi}_k(S_i) \tilde{\varphi}_k(S_j) \hat{\mu}(S_j)
\]

(45)

where \( \tilde{\varphi}_k(S_i) \) is given by
\[
\tilde{\varphi}_k(S_i) = \frac{\sum_{x \in S_i} \mu(x) \varphi_k(x)}{\sum_{x \in S_i} \mu(x)}
\]

(46)

In addition, when \( t = 1 \) Eq. 14 can also be expressed in terms of the edge matrix \( e \) and the degree \( d \) as
\[
E^* = \sum_{i,j=1}^{N} d(x)d(y) \left| \frac{e(x, y)}{d(x)d(y)} - \frac{\hat{e}^*(S_i, S_j)}{d(S_j)d(S_j)} \right|^2
\]

(47)

where
\[
\hat{e}^*(S_i, S_j) = \hat{e}^*(S_j, S_i) = \sum_{x \in S_i, y \in S_j} e(x, y)
\]

(48)

and
\[
\hat{d}(S_i) = \sum_{x \in S_i} d(x) = \sum_{j=1}^{N} \hat{e}^*(S_i, S_j)
\]

(49)

Eq. 48 gives the effective edge matrix on the reduced network on \( S \).
The cost of the $k$-means algorithm

To estimate the cost of the algorithm, we take another form of Eq. 33:

$$
\bar{E}(x, S_j) = \mu(x) \sum_{k=1}^{N} \sum_{y \in S_k} \mu(y) \left( \frac{p_t^2(x, y)}{\mu_2(y)} - 2p_t(x, y) \frac{\hat{p}^*(S_j, S_k)}{\mu(S_k)} + \frac{\hat{p}^{*2}(S_j, S_k)}{\tilde{\mu}^2(S_k)} \right)
$$

$$
= \sum_{k=1}^{N} \sum_{y \in S_k} \frac{\mu(x)}{\mu(y)} p_t^2(x, y) + \mu(x) \sum_{k=1}^{N} \frac{\hat{p}^{*2}(S_j, S_k)}{\tilde{\mu}(S_k)}
$$

$$
- 2\mu(x) \sum_{k=1}^{N} \sum_{y \in S_k} p_t(x, y) \frac{\hat{p}^*(S_j, S_k)}{\tilde{\mu}(S_k)}
$$

$$
\equiv P_1 + P_2 - P_3.
$$

(50)

Let $n$ denote the number of nodes, $m$ the number of links and $N$ the number of communities. The computation of $\{\hat{\mu}(S_k)\}_{k=1}^{N}$ and $\{\mu_k(x)\}_{k=1}^{N}$ are both $O(n)$. The computation of $\hat{p}^*(S_i, S_j)$ is $O(d_{ij})$, where $d_{ij}$ is the connecting degree from the $i$-th community to the $j$-th community. So the computation of $\{\hat{p}^*(S_i, S_j)\}_{i,j=1}^{N}$ is at most $O(m)$ (and it should actually be far less than $m$ because we do not need to take into account the connections within the communities).

For each fixed $x$ and $j$, the computation of $P_1$ is $O(d(x))$, where $d(x)$ is the degree of node $x$. The computation of $P_3$ is also $O(d(x))$. Note that the summation part in $P_2$ can be precomputed with $O(N^2)$ computational effort for all $j$, and as a result, the computation of $P_2$ for all $x$ and $j$ is $O(Nn + N^2)$. To compute $\bar{E}(x, S_j)$ for all $x$ and $j$, the computational effort is thus $O(N(2m + n) + N^2 + n + m)$. With the consideration that $N \ll n$ and $m$ is $O(n)$ for realistic networks, and we have that the computational cost per iteration step is $O(N(m + n))$.

If we take into account the average number of iterations $k_1$ and the number of trials $k_2$, we have the final computational effort $O(k_1 k_2 N(m + n))$. 