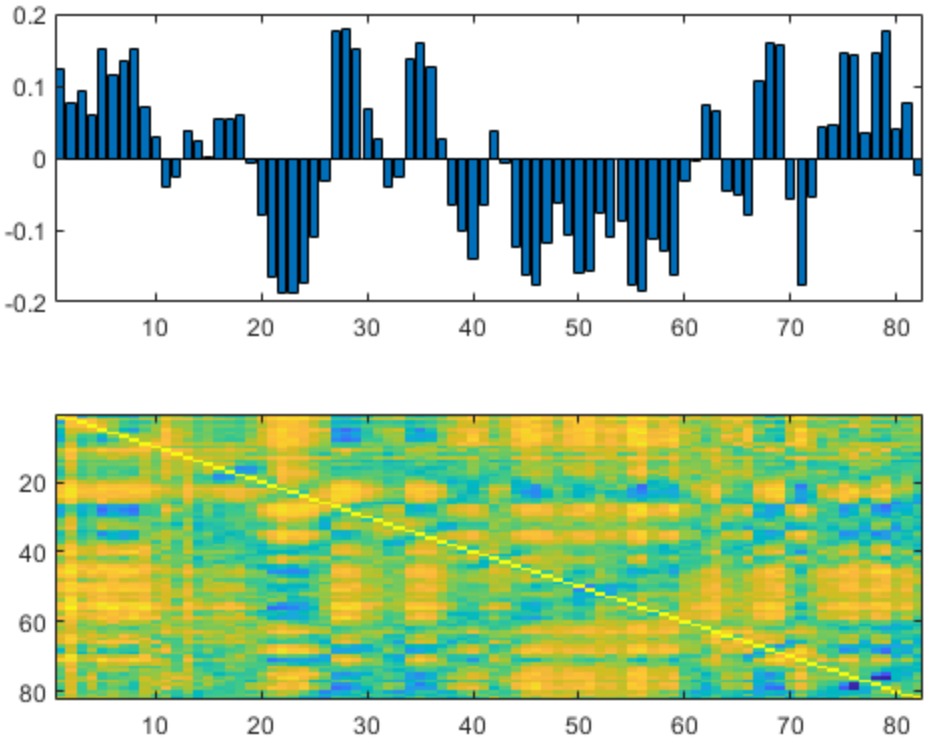
```
%name:Qingyuan Liu
load('PS4_data_v2/PS4_data_v2.mat')

%1(a)
CTCF_matrix = zeros(1,82);
for i = 1:length(regions_HiC_chr15)
    CTCF_matrix(i) = length(find(CTCF_chr15 > regions_HiC_chr15(1,i) &
 CTCF_chr15 < regions_HiC_chr15(2,i)));
end

CTCF_matrix = CTCF_matrix/norm(CTCF_matrix); %normalization

%1(a)
DNASE_matrix = zeros(1,82);
for i = 1:length(DNASEseq_chr15)
    if DNASEseq_chr15(i,2) < regions_HiC_chr15(2,82) &&
 DNASEseq_chr15(i,3) > regions_HiC_chr15(2,82)
        DNASE_matrix(82) = DNASE_matrix(82)+102*10e5-
DNASEseq_chr15(i,2);
        continue; % if the ending point exceeds the region length
 limit, calculate the length from lower end to limit
    end
    if DNASEseq_chr15(i,2) > regions_HiC_chr15(2,82)
        continue; % if the starting point exceeds the region length
 limit, skip
    end
    if fix(DNASEseq_chr15(i,2)/10e5) ~= fix(DNASEseq_chr15(i,3)/10e5)
        %if the starting point and the ending point at two different
        %regions, count them seperately

 DNASE_matrix(ceil(DNASEseq_chr15(i,2)/10e5)-20)=DNASE_matrix(ceil(DNASEseq_chr15(
 - DNASEseq_chr15(i,2);

 DNASE_matrix(ceil(DNASEseq_chr15(i,3)/10e5)-20)=DNASE_matrix(ceil(DNASEseq_chr15(
ceil(DNASEseq_chr15(i,2)/10e5)*10e5 + DNASEseq_chr15(i,3);

    else

 DNASE_matrix(ceil(DNASEseq_chr15(i,2)/10e5)-20)=DNASE_matrix(ceil(DNASEseq_chr15(
 DNASEseq_chr15(i,3) - DNASEseq_chr15(i,2);
    end

end
DNASE_matrix = DNASE_matrix/norm(DNASE_matrix);

%1(a)
RNA_matrix = zeros(1,82);
for i = 1:length(RNA_15)
    if RNA_15{i,2} < regions_HiC_chr15(2,82) && RNA_15{i,3} >
 regions_HiC_chr15(2,82)
        % if the ending point exceeds the region length limit,
 calculate the length from lower end to limit
```

```
        RNA_matrix(82) = RNA_matrix(82)+(102*10e5-
RNA_15{i,2})*RNA_15{i,4};
        continue;
    end
    if RNA_15{i,2} > regions_HiC_chr15(2,82)
       % if the starting point exceeds the region length limit, skip
       continue;
    end

    if fix(RNA_15{i,2}/10e5) ~= fix(RNA_15{i,3}/10e5)
         %if the starting point and the ending point at two different
        %regions, count them seperately
        RNA_matrix(ceil(RNA_15{i,2}/10e5)-20) =
 RNA_matrix(ceil(RNA_15{i,2}/10e5)-20) + (ceil(RNA_15{i,2}/10e5)*10e5
 - RNA_15{i,2})*RNA_15{i,4};
        RNA_matrix(ceil(RNA_15{i,3}/10e5)-20) =
 RNA_matrix(ceil(RNA_15{i,3}/10e5)-20) + (-ceil(RNA_15{i,2}/10e5)*10e5
 + RNA_15{i,3})*RNA_15{i,4};
    else
        %length * expression level
        RNA_matrix(ceil(RNA_15{i,2}/10e5)-20) =
 RNA_matrix(ceil(RNA_15{i,2}/10e5)-20) + (RNA_15{i,3} -
 RNA_15{i,2})*RNA_15{i,4};

    end

end
RNA_matrix=RNA_matrix / 10e5;
RNA_matrix = RNA_matrix/norm(RNA_matrix);

O_E_chr15 = raw_chr15./expected_chr15;
A = O_E_chr15;
D = diag(sum(O_E_chr15));
L = D - A;

L_sym = D^(-.5)*L*D^(-.5);      % normalized laplacian

[V,D] = eigs(L_sym,2,'smallestabs');
Fiedler_num = D(2,2);
Fiedler_vec = V(:,2);
```

```
Warning: First input matrix is close to singular or badly scaled.
 RCOND =
2.697448e-19. Results may be inaccurate.
```

```
figure, subplot(2,1,1), bar(Fiedler_vec)
xlim([0+.5 length(Fiedler_vec)+.5]) % sets the xlimits of the plot

subplot(2,1,2), imagesc(L_sym-diag(diag(L_sym)))    % image of
 normalized laplacian with diagonal removed
```




```
%We still need to construct correlation matrix as from Fiedler
 vectors, it
%clearly indicate some of the regions is more connected than others
 and we do
%not have clear idea if there is any correlation between regions. To
 find
%out which two regions are correlated, correlation matrix need to be
%calculated.

%calculate the Singular vector and fiedler vectors of the graph
%calculate the cross-entropy between two 1D features.
%[U,S,V] = svd(O_E_chr15);
%diff_1 = crossentropy(U(:,1)),DNASE_matrix);
%diff_2 = crossentropy(Fiedler_vec, DNASE_matrix);
```

Published with MATLAB® R2019b