Community discovery in social network: Applications, methods and emerging trends

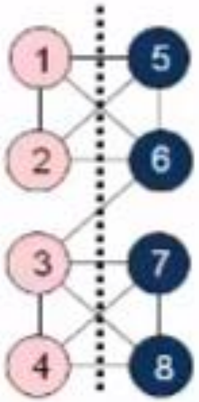
L18,19-SNA

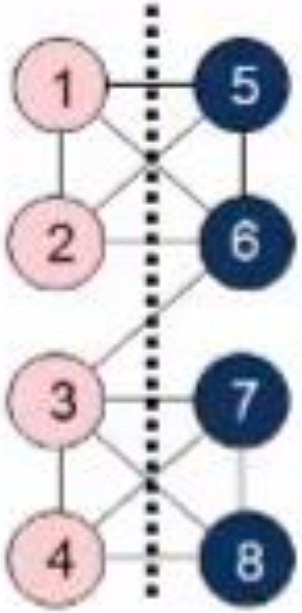
Partitioning using K-L algorithm: an example

Given: A graph with $2n$ nodes where each node has the same weight.

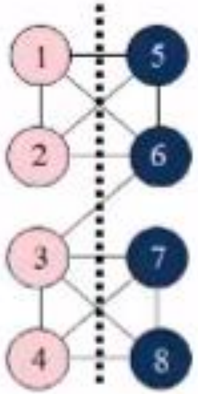
Goal: A partition (division) of the graph into two disjoint subsets A and B with minimum cut cost and $|A| = |B| = n$.

Example with n=4





Cut cost: 9
Not fixed:
1,2,3,4,5,6,7,8

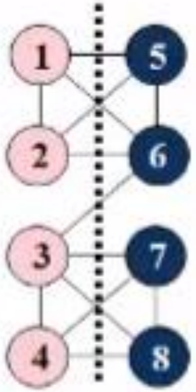


$D_1=E_1-I_1=2-1=1$

Cut cost: 9
Not fixed:
1,2,3,4,5,6,7,8

Benefit D_v of each node:

$D_1 = 1$	$D_5 =$
$D_2 =$	$D_6 =$
$D_3 =$	$D_7 =$
$D_4 =$	$D_8 =$



Compute gains for all possible swaps:

$g_{15}=D_1+D_5-2.c_{15} = 0$
$g_{25}=D_2+D_5-2.c_{25} =$
$g_{35}=D_3+D_5-2.c_{35} =$
$g_{45}=D_4+D_5-2.c_{45} =$

$g_{16}=D_1+D_6-2.c_{16} = 1$
$g_{26}=D_2+D_6-2.c_{26} =$
$g_{36}=D_3+D_6-2.c_{36} =$
$g_{46}=D_4+D_6-2.c_{46} =$

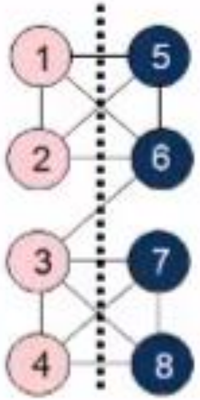
Cut cost: 9
Not fixed:
1,2,3,4,5,6,7,8

$g_{17}=D_1+D_7-2.c_{17} =$
$g_{27}=D_2+D_7-2.c_{27} =$
$g_{37}=D_3+D_7-2.c_{37} = 1$
$g_{47}=D_4+D_7-2.c_{47} =$

$g_{18}=D_1+D_8-2.c_{18} = 2$
$g_{28}=D_2+D_8-2.c_{28} =$
$g_{38}=D_3+D_8-2.c_{38} =$
$g_{48}=D_4+D_8-2.c_{48} =$

Benefit D_v of each node:

$D_1=1$	$D_5=1$
$D_2=1$	$D_6=2$
$D_3=2$	$D_7=1$
$D_4=1$	$D_8=1$



Cut cost: 9
Not fixed:
1,2,3,4,5,6,7,8

Compute gains for all possible swaps:

$g_{15}=D_1+D_5-2.c_{15}\ \ =0$
$g_{25}=D_2+D_5-2.c_{25}\ \ =0$
$g_{35}=D_3+D_5-2.c_{35}\ \ =3$
$g_{45}=D_4+D_5-2.c_{45}\ \ =2$

$g_{16}=D_1+D_6-2.c_{16}\ \ =1$
$g_{26}=D_2+D_6-2.c_{26}\ \ =1$
$g_{36}=D_3+D_6-2.c_{36}\ \ =2$
$g_{46}=D_4+D_6-2.c_{46}\ \ =3$

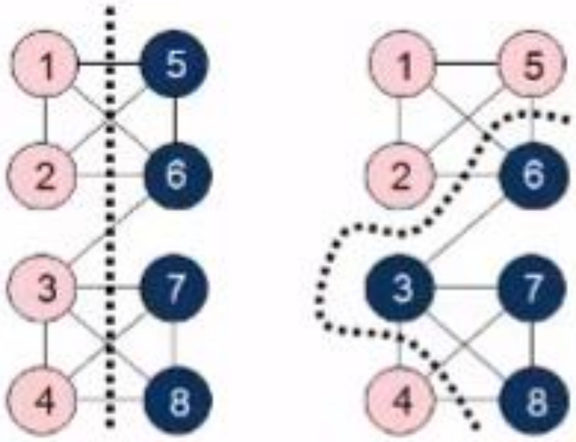
$g_{17}=D_1+D_7-2.c_{17}\ \ =2$
$g_{27}=D_2+D_7-2.c_{27}\ \ =2$
$g_{37}=D_3+D_7-2.c_{37}\ \ =1$
$g_{47}=D_4+D_7-2.c_{47}\ \ =0$

$g_{18}=D_1+D_8-2.c_{18}\ \ =2$
$g_{28}=D_2+D_8-2.c_{28}\ \ =2$
$g_{38}=D_3+D_8-2.c_{38}\ \ =1$
$g_{48}=D_4+D_8-2.c_{48}\ \ =0$

$g_1=2+1-0=3$
for **Swap (3,5)**

Swap(3,5) and Swap(4,6) lead to maximum gain of 3

Lets choose swap(3,5)

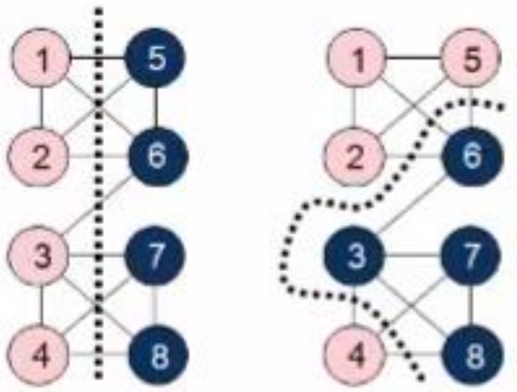


Cut cost: 9
Not fixed:
1,2,3,4,5,6,7,8

$D_1 = 1$	$D_5 = 1$
$D_2 = 1$	$D_6 = 2$
$D_3 = 2$	$D_7 = 1$
$D_4 = 1$	$D_8 = 1$

$g_1 = 2+1-0 = 3$

Swap (3,5)



If two elements $a \in A$ and $b \in B$ are interchanged, then the new D–values are given by

$$D'_x = D_x + 2c_{xa} - 2c_{xb}, \ \forall x \in A - \{a\}$$
$$D'_y = D_y + 2c_{yb} - 2c_{ya}, \ \forall y \in B - \{b\}$$

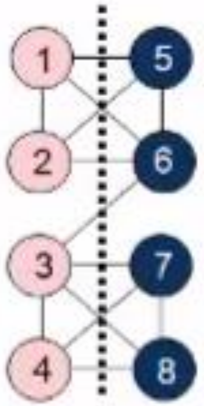
Cut cost: 9
Not fixed:
1,2,3,4,5,6,7,8

Cut cost: 6
Not fixed:
1,2,4,6,7,8

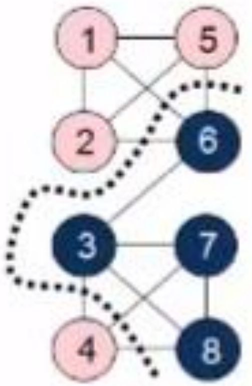
Update D values

$D_1 = 1$	$D_5 = 1$
$D_2 = 1$	$D_6 = 2$
$D_3 = 2$	$D_7 = 1$
$D_4 = 1$	$D_8 = 1$

$D_1' = -1$	$D_6' =$
$D_2' =$	$D_7' =$
$D_4' =$	$D_8' =$



Cut cost: 9
Not fixed:
1,2,3,4,5,6,7,8



Cut cost: 6
Not fixed:
1,2,4,6,7,8

$D_1 = 1$ $\quad D_5 = 1$
$D_2 = 1$ $\quad D_6 = 2$
$D_3 = 2$ $\quad D_7 = 1$
$D_4 = 1$ $\quad D_8 = 1$

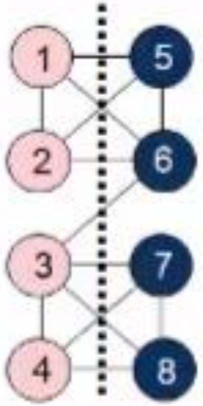
$D_1' = -1$ $\quad D_6' = 2$
$D_2' = -1$ $\quad D_7' = -1$
$D_4' = 3$ $\quad D_8' = -1$

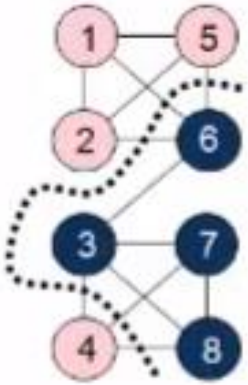
$A' = \{1,2,4\} \quad B' = \{6,7,8\}$

$g_{16} = D_1 + D_6 - 2c_{16} = -1$
$g_{17} = D_1 + D_7 - 2c_{17} =$
$g_{18} = D_1 + D_8 - 2c_{18} =$

$g_{26} = D_2 + D_6 - 2c_{26} =$
$g_{27} = D_2 + D_7 - 2c_{27} =$
$g_{28} = D_2 + D_8 - 2c_{28} =$

$g_{46} = D_4 + D_6 - 2c_{46} =$
$g_{47} = D_4 + D_7 - 2c_{47} =$
$g_{48} = D_4 + D_8 - 2c_{48} =$





Cut cost: 9
Not fixed:
1,2,3,4,5,6,7,8

Cut cost: 6
Not fixed:
1,2,4,6,7,8

$D_1 = 1$	$D_5 = 1$
$D_2 = 1$	$D_6 = 2$
$D_3 = 2$	$D_7 = 1$
$D_4 = 1$	$D_8 = 1$

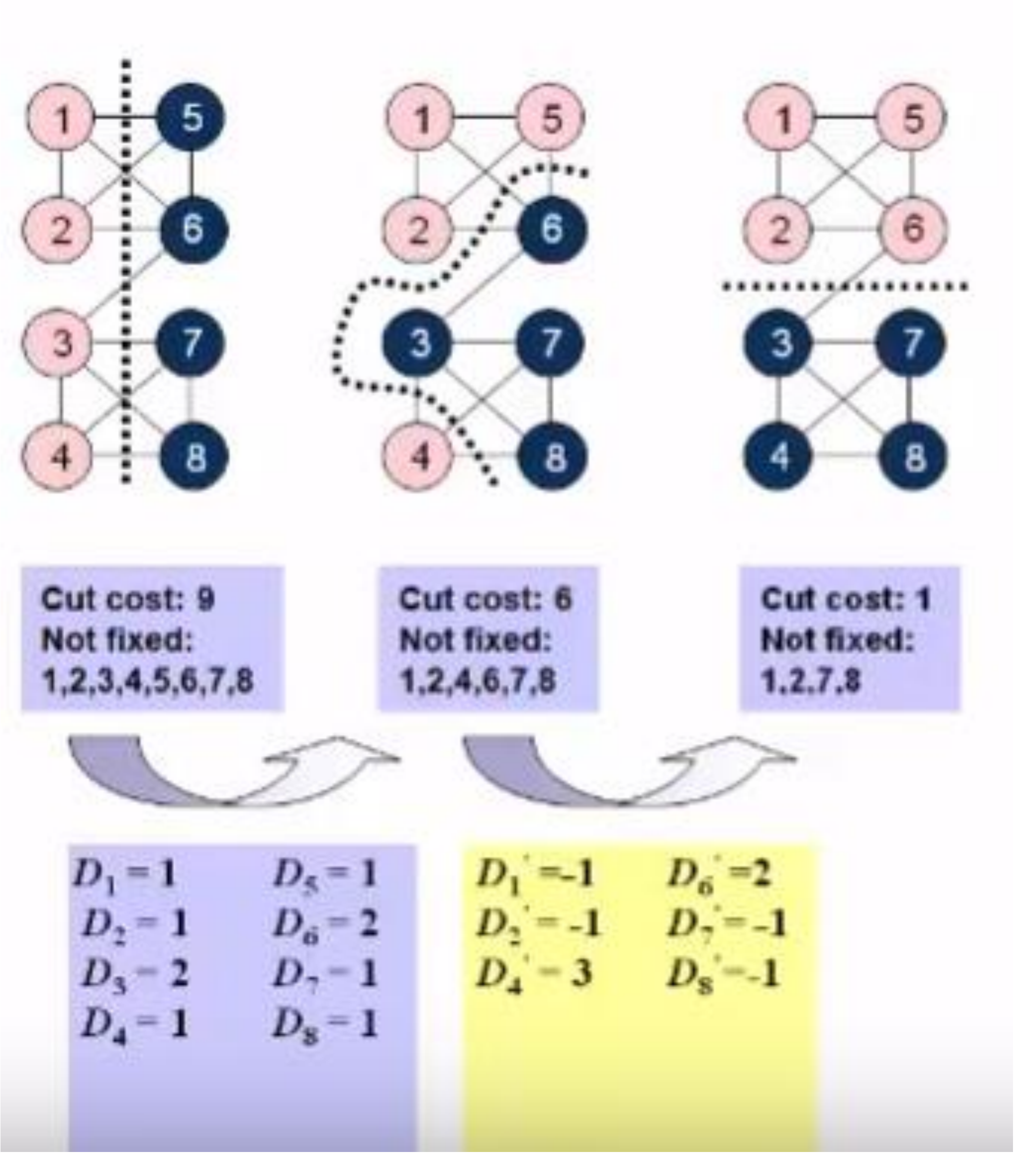
$D_1' = -1$	$D_6' = 2$
$D_2' = -1$	$D_7' = -1$
$D_4' = 3$	$D_8' = -1$

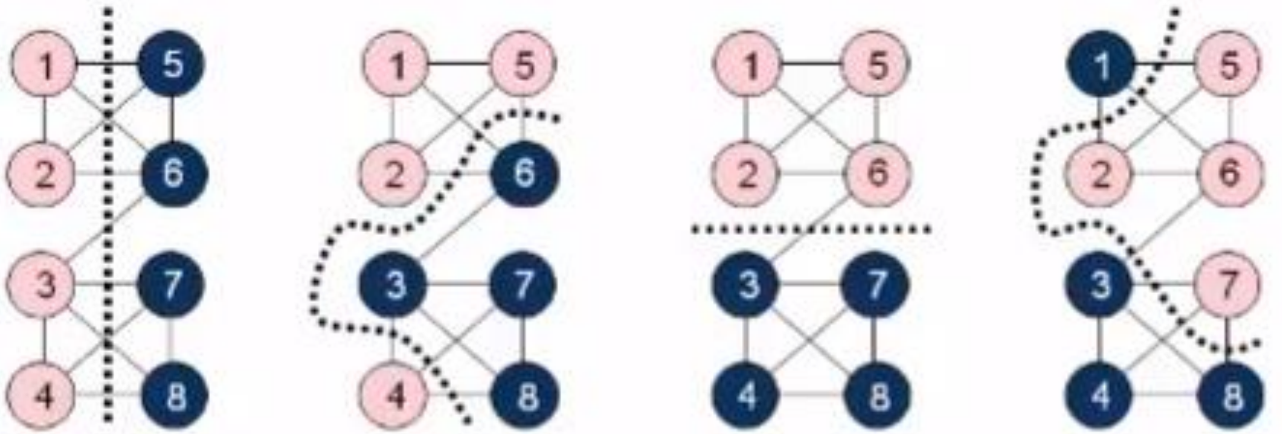
A'={1,2,4} B'={6,7,8}

$g_{16} = D_1 + D_6 - 2c_{16} = -1$
$g_{17} = D_1 + D_7 - 2c_{17} = -2$
$g_{18} = D_1 + D_8 - 2c_{18} = -2$

$g_{26} = D_2 + D_6 - 2c_{26} = -1$
$g_{27} = D_2 + D_7 - 2c_{27} = -2$
$g_{28} = D_2 + D_8 - 2c_{28} = -2$

$g_{46} = D_4 + D_6 - 2c_{46} = 5$
$g_{47} = D_4 + D_7 - 2c_{47} = 0$
$g_{48} = D_4 + D_8 - 2c_{48} = 0$


1
5
2
6
3
7
4
8
Cut cost: 9
Not fixed:
1,2,3,4,5,6,7,8
Cut cost: 6
Not fixed:
1,2,4,6,7,8
Cut cost: 1
Not fixed:
1,2,7,8
$D_1 = 1$ $D_5 = 1$
$D_2 = 1$ $D_6 = 2$
$D_3 = 2$ $D_7 = 1$
$D_4 = 1$ $D_8 = 1$
$D_1' = -1$ $D_6' = 2$
$D_2' = -1$ $D_7' = -1$
$D_4' = 3$ $D_8' = -1$



Cut cost: 9 **Not fixed:** **1,2,3,4,5,6,7,8**	**Cut cost: 6** **Not fixed:** **1,2,4,6,7,8**	**Cut cost: 1** **Not fixed:** **1,2,7,8**	**Cut cost: 7** **Not fixed:** **2,8**

$D_1 = 1 \quad D_5 = 1$

$D_2 = 1 \quad D_6 = 2$

$D_3 = 2 \quad D_7 = 1$

$D_4 = 1 \quad D_8 = 1$

$D_1' = -1 \quad D_6' = 2$

$D_2' = -1 \quad D_7' = -1$

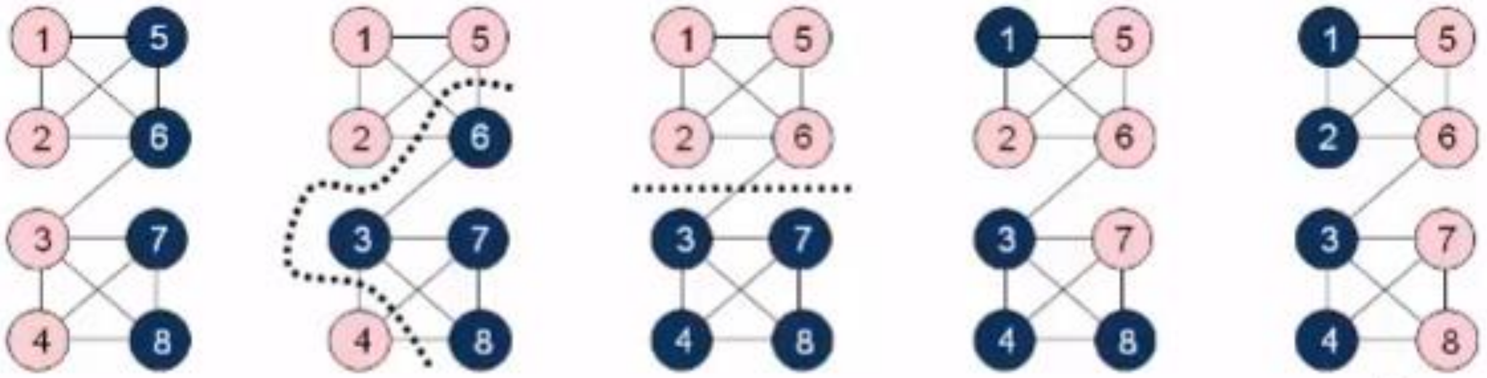
$D_4' = 3 \quad D_8' = -1$

$D_1'' = -3 \quad D_7'' = -3$

$D_2'' = -3 \quad D_8'' = -3$

$g_3 = -3 - 3 - 0 = -6$

for Swap (1,7)



Cut cost: 9
Not fixed:
1,2,3,4,5,6,7,8

Cut cost: 6
Not fixed:
1,2,4,6,7,8

Cut cost: 1
Not fixed:
1,2,7,8

Cut cost: 7
Not fixed:
2,8

Cut cost: 9
Not fixed:
–

$D_1 = 1$ $D_5 = 1$
$D_2 = 1$ $D_6 = 2$
$D_3 = 2$ $D_7 = 1$
$D_4 = 1$ $D_8 = 1$

$g_1 = 2+1-0 = 3$
for Swap (3,5)

$D_1' = -1$ $D_6' = 2$
$D_2' = -1$ $D_7' = -1$
$D_4' = 3$ $D_8' = -1$

$g_2 = 3+2-0 = 5$
for Swap (4,6)

$D_1'' = -3$ $D_7'' = -3$
$D_2'' = -3$ $D_8'' = -3$

$g_3 = -3-3-0 = -6$
for Swap (1,7)

$D_2''' = -1$
$D_8''' = -1$

$g_4 = -1-1-0 = -2$
for Swap (2,8)

$D_1 = 1$ $D_5 = 1$
$D_2 = 1$ $D_6 = 2$
$D_3 = 2$ $D_7 = 1$
$D_4 = 1$ $D_8 = 1$

$g_1 = 2+1-0 = 3$
for Swap (3,5)

$D_1' = -1$ $D_6' = 2$
$D_2' = -1$ $D_7' = -1$
$D_4' = 3$ $D_8' = -1$

$g_2 = 3+2-0 = 5$
for Swap (4,6)

$D_1'' = -3$ $D_7'' = -3$
$D_2'' = -3$ $D_8'' = -3$

$g_3 = -3-3-0 = -6$
for Swap (1,7)

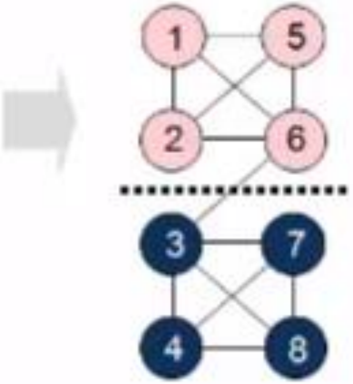
$D_2''' = -1$ $D_8''' = -$
1

$g_4 = -1-1-0 = -2$
for Swap (2,8)

Maximum positive gain $G_k = 8$ with $k = 2$.

Final Partition after single pass of KL algorithm

Since $G_k > 0$, the first $k = 2$ swaps after (3,5) and (4,6) are executed.



Iterative improvement- the partition obtained after the i^{th} pass becomes the initial partition of the (i+1) pass

Iterations are stopped when no further improvement is possible