1.

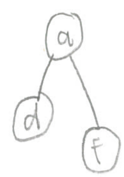
Solution: None of subgraphs have same substructure

: present in 3/4 subgraphs, not present in all

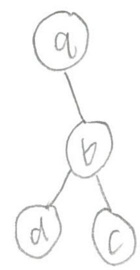
∴ labeled w/ +1 (not -1) b/c present in majority of decision stumps

An ensemble classifier would look at each subgraph's decision stump and if the subgraphs have the substructure for this instance, it receives a +1. Taken together, the 4× subgraph doesn't have the substructure.

new graph:



, since



isn't in there it's -1

+1 for other 3 subgraphs

Problem 2 Output

Summary motif results

====================

mfinder Version 1.20

MOTIF FINDER RESULTS:

Network name: problem2_input.txt

Network type: Directed

Num of Nodes: 280 Num of Edges: 2170

Num of Nodes with edges: 280

Maximal out degree (out-hub) : 39

Maximal in degree (in-hub) : 45

Roots num: 13 Leaves num: 21

Single Edges num: 1776 Mutual Edges num: 197

Motif size searched 4

Total number of 4-node subgraphs : 875316

Number of random networks generated : 10

Random networks generation method: Switches

Num of Switches range: 100.0-200.0, Success switches Ratio:0.762+-0.00

The following motifs were found:

Criteria taken : Nreal Zscore > 2.00

Pval ignored (due to small number of random networks)

Mfactor > 1.10

Uniqueness >= 4

Top Motifs List

MOTIF ID	NREAL	NRAND STATS	NREAL ZSCORE	NREAL PVAL	UNIQ VAL	CREAL [MILI]
204	2950	2192.8+-101.5	7.46	0.000	32	3.37

0 0 1 1

0 0 1 1

0 0 0 0

0 0 0 0

222	1606	73.1+-12.6	121.74	0.000	17	1.83

0 1 1 1

1 0 1 1

0 0 0 0

0 0 0 0

904	3027	1972.4+-61.2	17.24	0.000	27	3.46

0 0 0 1

0 0 0 1

1 1 0 0

0 0 0 0

906	1905	715.7+-79.0	15.05	0.000	22	2.18

0 1 0 1

0 0 0 1

1 1 0 0

0 0 0 0

922	2143	98.3+-34.2	59.79	0.000	18	2.45

0 1 0 1

1 0 0 1

1 1 0 0

0 0 0 0

2252	1447	584.4+-62.7	13.75	0.000	28	1.65

0 0 1 1

0 0 1 1

0 0 0 1

0 0 0 0

6552	803	60.9+-20.9	35.48	0.000	11	0.92

0 0 0 1

1 0 0 1

1 0 0 1

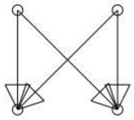
1 0 0 0

(Total No. of non-dangling motifs : 75)

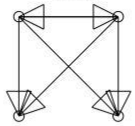
Full list includes 102 motifs

Top Motifs Graphs:

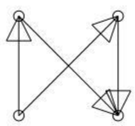
id204



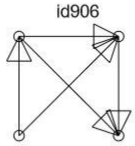
id222



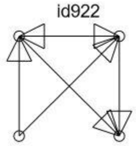
id904



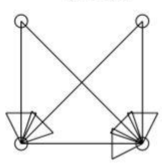
id906



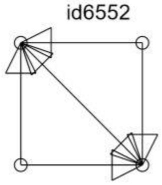
id922



id2252



id6552



3.

Solution:

similarities $\leq \alpha = 0.5$ (solid lines)
similarities $\geq \beta = 0.6$ (dashed lines)

Construct meta-graph $\Gamma(M)$ to measure sim. between graph patterns in M. Each node is a maximal subgraph. Edge exists between 2 nodes if their sim. is $\leq \alpha = 0.5$

Find maximal clique:

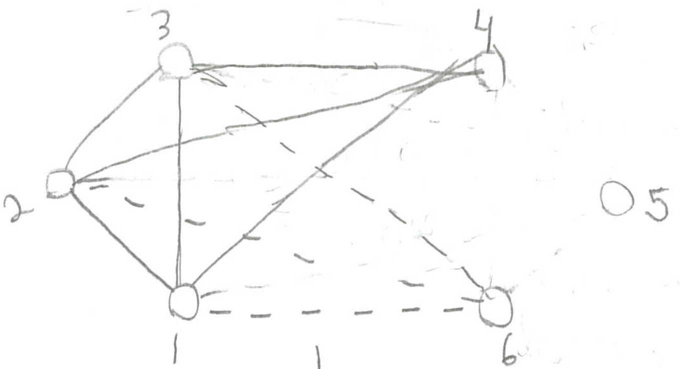
$R_1 = \{M_1, M_2, M_3, M_4\}$

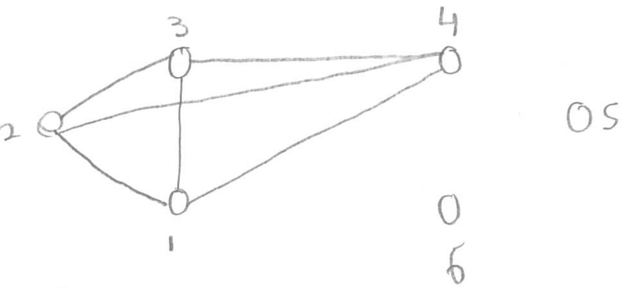
~~$R_2 = \{M_1, M_2, M_3\}$~~

~~$R_3 = \{M_1, M_2, M_4\}$~~

~~$R_4 = \{M_1, M_3, M_4\}$~~

R_2, R_3, R_4 not needed b/c R_1 captures them





$\gamma(R_1, D) = \{G \mid G \in D$ and there exists $G_a \in \{M_1, M_2, M_3, M_4\}, sim(G_a, G) \geq 0.6\}$
$= \{ M_6 \}$

Then $\Delta(R_1, D) = D - (R_1 \cup \gamma(R_1, D)) = \{M_5\}$ ← unrepresented by R_1

$\gamma(R_2, D) = \{G \mid G \in D$ and there exists $G_a \in \{M_1, M_2, M_3\}, sim(G_a, G) \geq 0.6\}$
$= \{ M_6 \}$

Then $\Delta(R_2, D) = D - (R_2 \cup \gamma(R_2, D)) = \{ M_4, M_5 \}$

$\gamma(R_3, D) = \{G \mid G \in D$ and there exists $G_a \in \{M_1, M_2, M_4\}, sim(G_a, G) \geq 0.6\}$
$= \{ M_6 \}$

Then $\Delta(R_3, D) = D - (R_3 \cup \gamma(R_3, D)) = \{ M_3, M_5 \}$

3.

$\gamma(R_4, D) = \{G \mid G \in D$ and there exists $G_a \in \{M_1, M_3, M_4\}, \ sim(G_a, G) \geq 0.6\}$

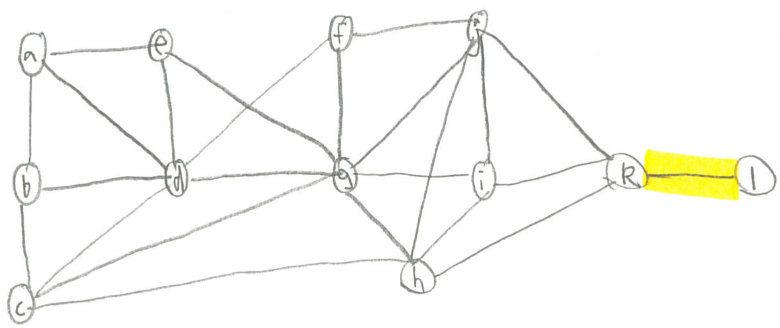
$= \{M_6\}$

Then $\Delta(R_4, D) = D - (R_4 \cup \gamma(R_4, D)) = \{M_2, M_5\}$

∴ The optimal α-orthogonal, β-representative set is R.

4.

Solution:



$k \geq 3$

3-truss

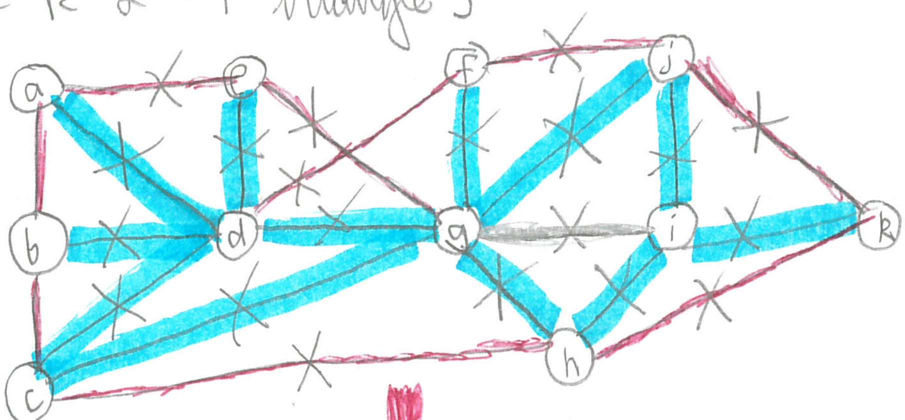
$k = 3$

$(k-2)$

Look for edges w/ support <1: (i,k) i.e., not involved in any triangles

Removing that edge doesn't cause support of any other edge to change to <1

Resulting graph is the 3-truss (every remaining edge is contained in $\geq k-2 = 1$ triangle)



$k = 4$

no 4-truss

all remaining edges would be removed

graph is empty, no more trusses

Edges w/ support <2: $(a,b)(b,c)$, (c,h), (a,e), (e,g), (d,f), (f,j), (h,k), (j,k) i.e., involved in only 1 triangle

Removing those edges causes support of some other edges to become <2: (a,d), (b,d), (c,d), (e,d), (c,a), (d,g), (g,f), (g,h), (g,j), (h,i), (j,i), (i,k)

Removal of those edges causes support of other edges to change to <2: (g,i)

5. a)

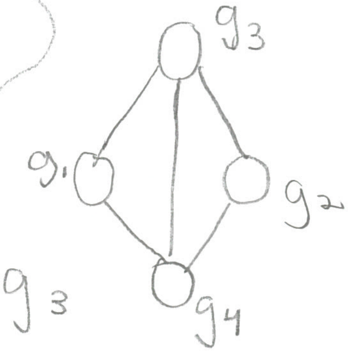
	degree	neighbor degree sequence
g_1	2	4, 4
g_2	6	4, 4, 3, 5, 4, 2
g_3	4	2, 6, 4, 2
g_4	4	2, 6, 4, 3
g_5	3	6, 4, 4
g_6	2	5, 4
g_7	5	6, 2, 4, 2, 2
g_8	4	6, 3, 2, 5
g_9	2	4, 5
g_{10}	2	6, 5

$g = g_1 \quad h = h_1 \quad f(h_1) = g_1 \Rightarrow f(h_2) = g_2 ; f(h_3) = g_3$

$h_1 < h_2, \; h_3 < h_4 ;$

$\checkmark \; g_1 < g_2, \; g_3 = g_4 \; \times$

none will work here



$f(h_1) = g_1 \Rightarrow f(h_2) = g_2 ;$

$f(h_3) = g_4 ; \; f(h_4) = g_3$

$\checkmark$

$g_1 < g_2, \; g_4 = g_3 \; \times$

$g = g_2 \quad h = h_1 \quad f(h_1) = g_2 \Rightarrow f(h_2) = g_6 ; \; f(h_3) = g_7$

$f(h_4) = g_8$

$h_1 < h_2, \; h_3 < h_4 ;$

$g_2 < g_6, \; g_7 < g_8 \; \times$

5. a)

$g = g_6 \quad h = h_1 \quad f(h_1) = g_6 \Rightarrow f(h_2) = g_2; \; f(h_3) = g_8$
$f(h_4) = g_7$

$h_1 < h_2, \; h_3 < h_4$

$g_6 < g_2, \; g_8 < g_7$ ✓ ✓

∴ only one that will work b/c $g_2 > g_6$ and $g_7 > g_8$

$g = g_4 \quad h = h_1 \quad f(h_1) = g_4 \Rightarrow f(h_2) = g_8; \; f(h_3) = g_2$
$f(h_4) = g_5$

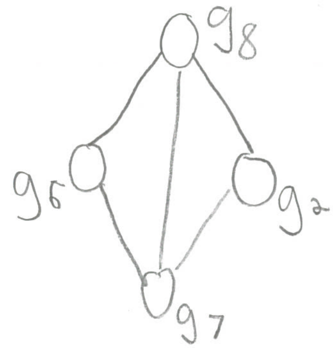
$h_1 < h_2, \; h_3 < h_4$

$g_4 = g_8, \; g_2 > g_5$ ✗ ✗

∴ none in this substructure will work because g_4 and g_8 have equal degrees

---

Instance found:



frequency of this pattern in the input graph is 1.

b) Zscore (S) = {frequency (S, G) - meanFrequency (S, G_R)}/σ
where σ = standard deviation

Zscore (S) = {1 - 2}/0.35
= -2.85714286

Pvalue (S) = (# random graphs in which S occurred more often than in

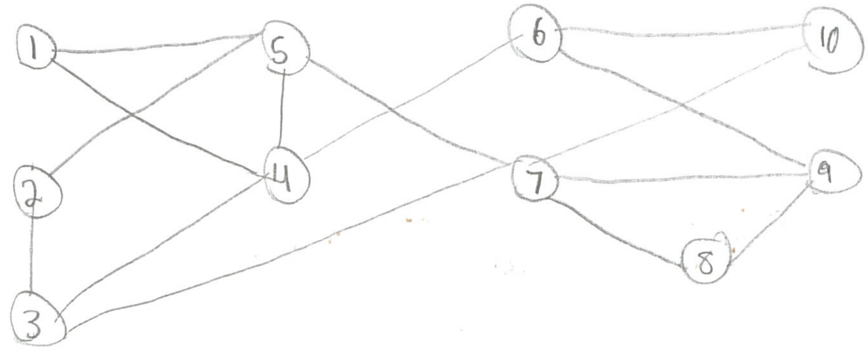
5.

Z-score < 2 → $|G_r| = 750 < 1000$

$-2.8571 < 2$

∴ H wouldn't be considered a motif of G. The calculated Z-score is significantly less than 2.

6.



d_i = # of neighbors of node v_i

R_i = # edges among these neighbors

C_i = clustering coefficient of v_i:

$$C_i = \begin{cases} \dfrac{R_i}{d_i \times (d_i - 1)/2} & d_i > 1 \\ 0 & d_i = 0 \text{ or } 1 \end{cases}$$

The higher the value of C_i, the more likely v_i and its neighbors resemble a clique

Let v_1 be the node labeled 1; its neighbors are 4 and 5

$d_1 = 2$

$R_1 = 1$

$C_1 = R_1/(d_1 \cdot (d_1 - 1)/2) = 1/(2 \cdot (2-1)/2) = 1/(2 \cdot \frac{1}{2}) = 1$

V_2's neighbors are 3 and 5

$d_2 = 2$

$R_2 = 0$

$C_2 = 0/(2 \cdot (2-1)/2) = 0$

V_3's neighbors are 2, 4, 7

$d_3 = 3$

$R_3 = 0$

$C_3 = 0/(3 \cdot (3-1)/2) = 0$

6.

V_4's neighbors are 3, 5, 6, and 8

$d_4 = 4$

$R_4 = 0$

$C_4 = 0/(4 \cdot (4-1)/2) = 0$

V_5's neighbors are 1, 2, 4, 7

$d_5 = 4$

$R_5 = 1$

$C_5 = 1/(4 \cdot (4-1)/2) = 1/(4 \cdot 3/2) = 1/6 = 0.1\overline{6}$

V_6's neighbors are 4, 9, 10

$d_6 = 3$

$R_6 = 0$

$C_6 = 0$

V_7's neighbors are 5, 8, 9, 10

$d_7 = 4$

$R_7 = 1$

$C_7 = 1/(4 \cdot (4-1)/2) = 1/(4 \cdot \frac{3}{2}) = 1/6 = 0.1\overline{6}$

V_8's neighbors are 4, 7, 9

$d_8 = 3$

$R_8 = 1$

$C_8 = 1/(3 \cdot (3-1)/2) = 1/(3 \cdot 1) = \frac{1}{3} = 0.3\overline{3}$

V_9's neighbors are 6, 7, 8

$d_9 = 3$

$R_9 = 1$

$C_9 = 1/(3 \cdot (3-1)/2) = 0.3\overline{3}$

V_{10}'s neighbors are 6, 7

$d_{10} = 2$

$$C = \sum_{i=1}^{n} C_i/n$$

$$= \frac{1+0+0+0+0.1\overline{6}+0+0.1\overline{6}+0.3\overline{3}+0.3\overline{3}+0}{10}$$

$$= \underline{0.20}$$

Since the community structure value is low, the nodes are loosely clustered into groups in the network

7.

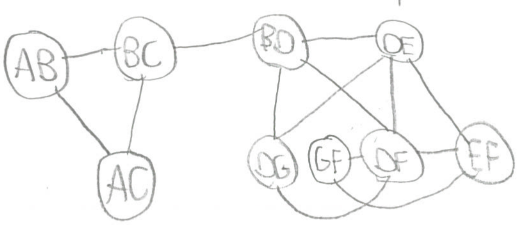
a) The interpretation of the edges of the resulting graph is that if Friend X (node X) and Friend Y (node Y) are related in some way through their connecting edge, an edge connecting Friend X and Friend Z also means there will be a relationship in same way for Friend Y and Friend Z through an edge by both of their relationships to Friend X.

b) G' - representing this

<u>edges in G</u>

A-B, A-C, B-C, B-D, D-E, D-F, D-G, E-F, G-F

these will be node pairings in G'

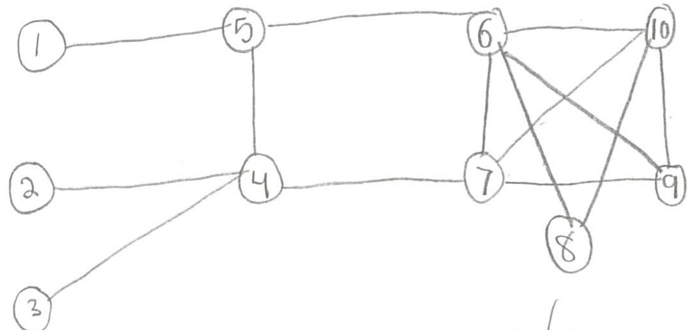


The relationships containing a common letter will all be linked w/ an edge

c) The degree of a node XY in G' compared to the degrees of X and Y in G: up to a certain point, the connections will all relate but as the number of people w/ relationships in the community grows, the network starts to break up into isolated, disconnected components and/or fewer people in each connected component.

The degrees of a node XY in G' should be the same as the degrees X and Y in G. The nodes AB, BC, and AC are a good example to the nodes A, B, and C.

8.



$V_S = \{6, 7, 8, 9, 10\}$

$V_S = \{6, 7, 9, 10\}$

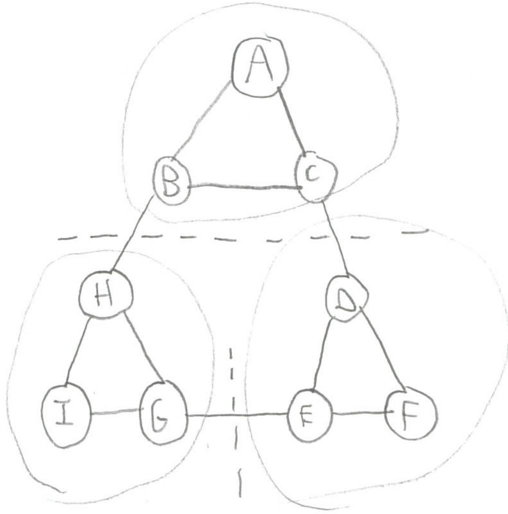
$|E_S| / (|V_S| \cdot (|V_S| - 1)/2) = 8/(5 \cdot 4/2)$
$= 0.8 > 0.65$

$|E_S| / (|V_S| \cdot (|V_S| - 1)/2) = 6/(6 \cdot 5/2)$
$= 0.4 < 0.65$

All other cliques would generate smaller values for their γ-density b/c they will have less edges relative to the number of nodes that are applied to the denominator of the γ-density calculation.

$\boxed{V_S = \{6, 7, 8, 9, 10\} \text{ is a clique}}$

9.



2 cuts

3 edges removed

3 communities created of equal size

Let S be the set of nodes in subgraph S

$Vol(S)$ = # edges w/ at least one end in S

$Cut(S, T, U)$ = # edges that connect a node in S to a node in $T + U$

$$NormalizedCut(S, T, U) = \underline{(Cut(S, T, U) / Vol(S)) + (Cut(S, T, U) / Vol(T)) + (Cut(S, T, U) / Vol(U))}$$

minimize this

① $S = \{A, B, C\}$

$T = \{G, H, I\}$

$U = \{D, E, F\}$

$Vol(S) = 5$, $Cut(S, T, U) = 3$

$Vol(T) = 5$, $Cut(S, T, U) = 3$

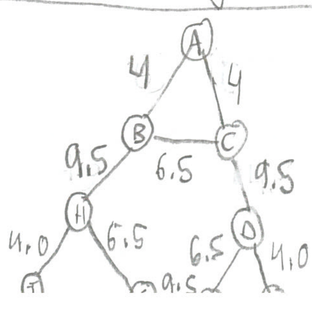
$Vol(U) = 5$, $Cut(S, T, U) = 3$

$NormalizedCuts(S, T, U) = (3/5) + (3/5) + (3/5) = 1.8$

WRONG METHOD

edge betweenness algorithm - Right Method

See R Code



	A	B	C	D	E	F	G	H	I
A	0	1	1	0	0	0	0	0	0
B	1	0	1	0	0	0	0	1	0
C	1	1	0	1	0	0	0	0	0
D	0	0	1	0	1	1	0	0	0
E	0	0	0	1	0	1	1	0	0
F	0	0	0	1	1	0	0	0	0
G	0	0	0	0	1	0	0	1	1
H	0	1	0	0	0	0	1	0	1
I	0	0	0	0	0	0	1	1	0

```
> edges=c(c(0,1,1,0,0,0,0,0,0), c(1,0,1,0,0,0,0,1,0), c(1,1,0,1,0,0,0,0,0), c(0,0,1,0,1,1,0,0,0), c(0,0,0,1,0,1,1,0,0), c(0,0,0,1,1,0,0,0,0), c(0,0,0,0,1,0,0,1,1), c(0,1,0,0,0,0,1,0,1), c(0,0,0,0,0,0,1,1,0))

> nodeNames=c("A","B","C","D","E","F","G","H","I")

> adjMatrix=matrix(edges,nrow=9,ncol=9,byrow=TRUE,dimnames=list(nodeNames,nodeNames))

>

>

>

>

>

>

> adjMatrix

  A B C D E F G H I

A 0 1 1 0 0 0 0 0 0

B 1 0 1 0 0 0 0 1 0

C 1 1 0 1 0 0 0 0 0

D 0 0 1 0 1 1 0 0 0

E 0 0 0 1 0 1 1 0 0

F 0 0 0 1 1 0 0 0 0

G 0 0 0 0 1 0 0 1 1

H 0 1 0 0 0 0 1 0 1

I 0 0 0 0 0 0 1 1 0

> library("igraph")


Attaching package: 'igraph'


The following objects are masked from 'package:stats':


    decompose, spectrum
```

The following object is masked from ‘package:base’:

    union

Warning message:

package ‘igraph’ was built under R version 3.6.3

```
> g=graph_from_adjacency_matrix(adjMatrix, mode="undirected")
> g
IGRAPH 1918485 UN-- 9 12 --
+ attr: name (v/c)
+ edges from 1918485 (vertex names):
 [1] A--B A--C B--C B--H C--D D--E D--F E--F E--G G--H G--I H--I
> edge_betweenness(g)
 [1] 4.0 4.0 6.5 9.5 9.5 6.5 4.0 4.0 9.5 6.5 4.0 4.0
> cluster_edge_betweenness(g,weights=NULL,directed=FALSE)
IGRAPH clustering edge betweenness, groups: 3, mod: 0.42
+ groups:
  $`1`
  [1] "A" "B" "C"

  $`2`
  [1] "D" "E" "F"

  $`3`
  [1] "G" "H" "I"
```