Gokhulnath Thirumaran
675086474

1) a) Given: The web has no ends

To prove: $w(r') = w(r)$.

Proof:

$$w(r') = \sum_{i=1}^{n} r'_i$$

$$= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} M_{ij} r_j \right)$$

$$= \sum_{j=1}^{n} \left(\sum_{i=1}^{n} M_{ij} r_j \right)$$

$$= \sum_{j=1}^{n} r_j \left(\sum_{i=1}^{n} M_{ij} \right)$$

$$= \sum_{j=1}^{n} r_j$$

$$= w(r)$$

Since $\sum_{i=1}^{n} M_{ij} = 1$

As, sum of jth column of Matrix M on the left side. Since the webs has no ends k_j values equal to $1/k_j$ in the j^{th} column. So the sum of each column is therefore equal to 1.

1) b) We teleport to a random node with probability $1-\beta$, where $0 < \beta < 1$. So,

$$r_i' = \beta\left(\sum_{j=1}^{n} M_{ij} r_j\right) + \frac{1-\beta}{n}$$

To determine when $w(r') = w(r)$ is true,

$$\sum_{i=1}^{n} r_i = \sum_{i=1}^{n} r_i'$$

$$\sum_{i=1}^{n} r_i = \sum_{i=1}^{n}\left(\beta\left(\sum_{j=1}^{n} M_{ij} r_j\right) + \frac{1-\beta}{n}\right)$$

$$\sum_{i=1}^{n} r_i = \beta \sum_{j=1}^{n}\left(\sum_{j=1}^{n} M_{ij} r_j\right) + \sum_{i=1}^{n} \frac{1-\beta}{n}$$

$$\sum_{i=1}^{n} r_i = \beta \sum_{j=1}^{n}\left(\sum_{i=1}^{n} M_{ij}\right) + 1 - \beta$$

$$\sum_{i=1}^{n} r_i = \beta \sum_{j=1}^{n} r_j + 1 - \beta$$

$$w(r) = \beta w(r) + 1 - \beta$$

$$w(r) - \beta w(r) = 1 - \beta$$

$$w(r)(1-\beta) = 1 - \beta$$

$$w(r) = 1$$

Thus $w(r') = w(r)$ holds true, if $w(r) = 1$

1) c) At each iteration, we teleport to nodes with probability $1-\beta$ and from dead nodes with probability 1. We choose randomly node uniformly to teleport to. Assume D as set of dead nodes.

$$r_i' = \underbrace{\beta\left(\sum_{j \notin D} M_{ij} r_j\right)}_{1} + \underbrace{\frac{1-\beta}{n} \sum_{j \notin D} r_j}_{2} + \underbrace{\frac{1}{n} \sum_{j \in D} r_j}_{3}.$$

1 ⇒ with probability β a web surfer chooses an out-link on their current page

2 ⇒ $\frac{1-\beta}{n}$ represents that a web surfer opens a new page

3 ⇒ teleportation from a dead node.

$$\Rightarrow \quad w(r') = \sum_{i=1}^{n} \left(\beta \sum_{j \notin D} M_{ij} r_j + \frac{1-\beta}{n} \sum_{j \notin D} r_j + \frac{1}{n} \sum_{j \in D} r_j\right).$$

given $w(r) = \sum_{j=1}^{n} r_j = 1$ and $\sum_{i=1}^{n} M_{ij} = 1$

$$w(r') = \sum_{i=1}^{n} \left(\beta \sum_{j \in D} M_{ij} r_j + \frac{1-\beta}{n} \sum_{j \notin D} r_j + \frac{1}{n} \sum_{j \in D} r_j\right)$$

$$= \beta \sum_{j \notin D} \left(\sum_{i=1}^{n} M_{ij} r_j\right) + \frac{1-\beta}{n} \sum_{j \notin D} r_j \left(\sum_{i=1}^{n} 1\right) + \frac{1}{n} \sum_{j \in D} r_j \left(\sum_{i=1}^{n} 1\right)$$

4

$$= \beta \sum_{j \notin D} r_j + (1-\beta) \sum_{j \notin D} r_j + \sum_{j \in D} r_j$$

$$= \sum_{j \notin D} r_j + \sum_{j \in D} r_j$$

$$= \sum_{j=1}^{4} r_j$$

$$= 1$$

2) a)

```
The top 5 node ids with the PageRank scores:
Node id: 263, PageRank score: 0.002020291181518219
Node id: 537, PageRank score: 0.00194334157145315
Node id: 965, PageRank score: 0.0019254478071662631
Node id: 243, PageRank score: 0.0018526340162417312
Node id: 285, PageRank score: 0.0018273721700645142
```

```
The bottom 5 node ids with the PageRank scores:
Node id: 558, PageRank score: 0.0003286018525215297
Node id: 93, PageRank score: 0.00035135689375l6577
Node id: 62, PageRank score: 0.00035314810510596274
Node id: 424, PageRank score: 0.00035481538649301454
Node id: 408, PageRank score: 0.00038779848719291705
```

2) b)

```
The 5 node ids with the highest hubbiness scores:
Node id: 840, hubbiness score: 1.0
Node id: 155, hubbiness score: 0.9499618624906543
Node id: 234, hubbiness score: 0.8986645288972264
Node id: 389, hubbiness score: 0.863417110184379
Node id: 472, hubbiness score: 0.8632841092495217
```

```
The 5 node ids with the lowest hubbiness scores:
Node id: 23, hubbiness score: 0.042066854890936534
Node id: 835, hubbiness score: 0.05779059354433016
Node id: 141, hubbiness score: 0.06453117646225179
Node id: 539, hubbiness score: 0.06602659373418492
Node id: 889, hubbiness score: 0.07678413939216454
```

```
The 5 node ids with the highest authority scores:
Node id: 893, hubbiness score: 1.0
Node id: 16, hubbiness score: 0.9635572849634398
Node id: 799, hubbiness score: 0.9510158161074016
Node id: 146, hubbiness score: 0.9246703586198444
Node id: 473, hubbiness score: 0.899866197360405
```

```
The 5 node ids with the lowest authority scores:
Node id: 19, hubbiness score: 0.05608316377607618
Node id: 135, hubbiness score: 0.06653910487622794
Node id: 462, hubbiness score: 0.07544228624641902
Node id: 24, hubbiness score: 0.08171239406816946
Node id: 910, hubbiness score: 0.08571673456144878
```

3) a) To prove: C_i is a clique for any $i > 1$

Proof: C_i is defined as a set of nodes of G that are divisible by i, $i > 0$. So every pair of nodes has a common factor i and are connected.

This implies that there is a edge between every two nodes in C_i, so C_i is clique.

3) b) C_i is a maximal clique if only if i is a prime.no.

If i is a prime, then all no. between 2 and 1000000 that are divisible by i are in C_i. If we add a node v to C_i, the added node v is not divisible by i, since all numbers divisible by i are already in C_i. So v will not connect to all the nodes already in C_i and $C_i \cup \{v\}$ will not have the property of clique. So, C_i really is a maximal clique and that condition that is prime is sufficient.

If i is not prime, we can write it as $i = p_1^{k_1} \cdots p_u^{k_u}$. 7
A node p_1 is not in C_i since it is not divisible by i.
But if we add it in C_i it will connect to all the nodes already in C_i because i and p_1 have a common p_1 factor, thus forming a clique.
So, for i not prime, C_i is not a maximal clique.

3) c) C_2 is the largest clique among all the clique of form C_i, so cardinality of a clique in this form is

$$|C_i| = \left\lfloor \frac{1000000}{i} \right\rfloor$$

Since $i=2$ is the smallest number, $|C_2| = 500000$ is the highest of all C_i, $i > 1$

Clique that are not form of C_i are sets of nodes that all have some common factor, but not all nodes that are divisible by this common factor are necessary in this clique. So cliques of this form are subsets of C_i for some i and therefore have smaller cardinality than C_2.

Thus C_2 is the largest clique.

4| a) i) To prove:

$$|A(S)| \geq \frac{\epsilon}{1+\epsilon}|S|$$

Proof:

Since $\overline{A(S)} = \{i \in S \mid deg_S(i) > 2(1+\epsilon)\rho(S)\}$

so, $|\overline{A(S)}| = |S| - |A(S)|$ —— ①

w.r.t. sum of all deg. in a graph $= 2|E(S)|$
as every edge is counted twice

$\overline{A(S)}$ is a subgraph of S, So sum of all degrees of vertices in $\overline{A(S)}$ is at most sum of all degrees of vertices in S.

So,

$$2|E(S)| \leq \sum_{i \in S} deg_S(i) \leq \sum_{i \in \overline{A(S)}} deg_S(i)$$ —— ②

Now sum of all the degree of vertices in $\overline{A(S)}$.

$$\sum_{i \in \overline{A(S)}} deg_S(i) > \sum_{i \in \overline{A(S)}} 2(1+\epsilon)\rho(S) = |\overline{A(S)}| \cdot 2(1+\epsilon)\rho(S)$$

from ②

$$|\overline{A(S)}| \cdot 2(1+\epsilon)\rho(S) < 2|E(S)|$$

from (1)

$$(|S| - |A(S)|) \cdot 2(1+\epsilon)\rho(S) < 2\rho(S)|S|$$

$$|S| - |A(S)| < \frac{|S|}{1+\epsilon}$$

$$|S| \cdot \left(1 - \frac{1}{1+\epsilon}\right) < |A(S)|$$

$$|S| \cdot \left(\frac{\epsilon}{1+\epsilon}\right) < |A(S)|$$

Thus proved.

ii) To prove:

Algorithm terminate at $O(\log_{1+\epsilon}(n))$ iteration, where $|S| = n$.

Proof:

Let's denote, S_i as a subgraph obtained in ith iteration. Its cardinality

$$|S_i| = |S_{i-1}| - |A(S_{i-1})| \leq |S_{i-1}| - \frac{\epsilon}{1+\epsilon}|S_{i-1}| \leq |S_{i-1}| \cdot \left(\frac{1}{1+\epsilon}\right)$$

Cardinality of S at the beginning is $|S_0| = n$. after k iteration S_k has cardinality ~~S_k~~

$$S_k = n \cdot \left(\frac{1}{1+\epsilon}\right)^k.$$

we need to find highest k, for which $|S_k|$ is still nonzero.

So,

$$0 < |S_k| \leq n \cdot \left(\frac{1}{1+\epsilon}\right)^k$$

$$1 \leq n \cdot \left(\frac{1}{1+\epsilon}\right)^k$$

$$\frac{1}{n} \leq (1+\epsilon)^{-k}$$

$$\log_{1+\epsilon}\left(\frac{1}{n}\right) \leq -k$$

$$-\log_{1+\epsilon}(n) \leq -k$$

$$\log_{1+\epsilon}(n) \geq k$$

Thus its proved that the alogorithm takes at most $\log_{1+\epsilon}(n)$ steps.

4) b) i) To prove: ~~$|E(S)| \rho \geq \frac{\epsilon}{1+\epsilon} |S|$~~

$$\deg_{S'}(v) \geq \rho^*(G).$$

Proof:

$$\rho(S) = \frac{|E(S)|}{|S|}$$

$\rho(S^*)$ is the highest among all densities of subgraph,

it has to include all possible edges between nodes in S^* that are in G.

So

$$\rho^*(G) = \rho(S^*) \quad — (1)$$

There exists vertex $v \in S^+$

$$\deg_{S^*}(v) < \rho^*(G) \quad — (2)$$

Since $\bar{S} = S^* \setminus \{v\}$, so,

$$\rho(S^*) = \frac{E[S^*]}{|S^*|}$$

$$= \frac{E[\bar{S}] + \deg_{S^*}(v)}{|S^*|}$$

$$= \frac{|S^*|-1}{|S^*|} \frac{E[\bar{S}]}{|S^*|-1} + \frac{\deg_{S^*}(v)}{|S^*|}$$

$$\rho(S^*) = \left(1 - \frac{1}{|S^*|}\right)\rho(\bar{S}) + \frac{1}{|S^*|}\deg_{S^*}(v)$$

From (1) & (2) it follows, that

$$\deg_{S^*}(v) < \rho^*(G) = \rho(S^*)$$

So,

$$\rho(\bar{S}) \geq \rho(S^*).$$

But this is a contradiction with the fact that S^* is the densest subgraph.

ii) To prove:

$$2(1+\epsilon)\rho(S) \geq \rho^*(G)$$

Proof: Assuming there exist a node $v \in S^* \cap A(S)$, we can prove this

As $v \in A(S)$, from 4 (a)

$$\deg_S(v) \leq 2(1+\epsilon)\rho(S).$$

As $v \in S^*$, from 4 (b) (i)

$$\rho^*(G) \geq \deg_{S^*}(v)$$

From S^*, we know that $S^* \subset S$, so each node in S^* has smaller degree than the same node in S: $\deg_{S^*}(v) \leq \deg_S(v)$

So,

$$2(1+\epsilon)\rho(S) \geq \deg_S(v) \geq \deg_{S^*}(v) \geq \rho^*(G)$$

iii) To prove:

$$\rho(\tilde{S}) \geq \frac{1}{2(1+\epsilon)} \rho^{*}(G)$$

Proof: In every iteration we remove all the nodes from $A(S)$ if $\rho(S) > \rho(\tilde{S})$. From some step onwards $\rho(S) \leq \rho(\tilde{S})$ will be true. Therefore $\tilde{S} \leftarrow S$ will never again be executed. While $\rho(S)$ will become smaller with each iteration $\rho(\tilde{S})$ will stay same.

In final iteration we get $\rho(S) \leq \rho(\tilde{S})$

From 4(b)(ii)

$$\rho^{*}(G) \leq 2(1+\epsilon)\,\rho(S) \leq 2(1+\epsilon)\,\rho(\tilde{S})$$

Thus proved.

5) a)

```
{'doc1': [-201.49, 35.3, -38.34, 19.77, 8.59, 18.24, 18.29, 22.89, 19.29],
 'doc2': [-147.26, -70.19, -26.35, 5.38, 54.18, -2.28, 2.21, 10.97, -8.43],
 'doc3': [-145.09, 26.81, 38.77, 33.8, 3.94, -36.21, 22.54, -44.29, -4.61],
 'doc4': [-190.2, -43.32, 13.14, -15.7, -39.66, 1.77, 14.61, 24.71, -5.47],
 'doc5': [-132.83, -19.84, 57.46, -16.56, 0.53, 5.76, 29.67, 11.23, -2.67],
 'doc6': [-122.42, 87.1, -0.45, -64.67, 6.6, -33.02, -25.73, 17.98, -19.92],
 'doc7': [-176.57, -24.33, 26.83, -34.24, 30.45, 24.33, 2.54, -13.11, -4.95],
 'doc8': [-111.91, 10.42, -52.93, -1.56, -5.22, 57.57, -19.43, -36.29, -13.46],
 'doc9': [-181.56, -71.17, -15.71, -27.35, -43.39, -3.15, -3.69, -6.77, -5.0],
 'doc10': [-206.53, -10.93, 9.4, -33.49, 3.95, -30.78, -17.87, -25.97, 27.13],
 'doc11': [-169.18, 31.94, 19.64, 27.51, 42.18, 8.29, 5.42, 8.96, -6.43],
 'doc12': [-124.33, -37.22, -60.29, 61.51, -2.53, -43.56, -21.63, 7.94, -8.48],
 'doc13': [-105.75, 28.98, -54.54, -22.75, 4.85, 6.57, -4.21, 6.63, 18.69],
 'doc14': [-185.75, 71.43, -19.03, 29.17, -37.05, 6.43, 22.3, -1.2, -6.0],
 'doc15': [-134.84, 5.53, 85.12, 46.08, -13.35, 23.5, -49.63, 13.29, 6.1]}
```

5) b).

15

```
{'word1': [-198.48, -72.47, 0.17, -54.0, 2.16, 42.5, 4.41, 10.14, 0.93],
 'word2': [-229.25, -36.06, 4.4, 61.75, 52.02, -24.49, 5.65, 22.33, 22.12],
 'word3': [-169.56, 25.05, 98.28, -32.86, 19.21, -24.56, -48.48, -5.29, -7.1],
 'word4': [-184.76, 15.79, 25.26, -39.47, -48.4, 1.3, 12.88, 40.16, 15.81],
 'word5': [-169.09, 4.15, -32.55, 67.41, -59.41, 3.88, -38.9, 6.32, -4.37],
 'word6': [-173.17, -72.72, 4.26, -9.66, -31.15, -61.04, 26.8, -32.39, -6.43],
 'word7': [-169.51, 110.36, -45.99, -28.28, 0.71, -10.0, 4.62, -29.09, 21.04],
 'word8': [-240.77,
  -30.04,
  -89.27,
  -25.31,
  26.59,
  13.63,
  -23.31,
  -9.46,
  -9.86],
 'word9': [-224.5, 18.01, 62.67, 43.26, -1.27, 54.23, 24.96, -31.9, -4.28],
 'word10': [-169.95, 69.79, -14.04, 5.15, 14.9, -16.15, 28.44, 30.92, -31.49]}
```