1 (a)

$$w(r') = \sum_i r_i' = \sum_i \sum_j M_{ij} r_j = \sum_j r_j \sum_i M_{ij} = \sum_j r_j 1 = w(r)$$

1 (b)

$$\boxed{w(r) = 1}$$

$$w(r') = \sum_i r_i' = \sum_i \left(\beta \sum_j M_{ij} r_j + (1-\beta)\frac{1}{n} \right)$$

$$= 1 - \beta + \beta \sum_j r_j \sum_i M_{ij} = 1 - \beta + \beta \sum_j r_j = 1 - \beta + \beta w(r) = 1$$

1 (c)

$$\boxed{r_i^{\acute{}} = \beta \sum_j M_{ij} r_j + \sum_{j \notin D} (1-\beta) r_j \frac{1}{n} + \sum_{j \in D} r_j \frac{1}{n} = \beta \sum_j M_{ij} r_j + \frac{(1-\beta)}{n} + \frac{\beta}{n} \sum_{j \in D} r_j}$$

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

$$= \beta \sum_{j \notin D} r_j + \sum_i \left(\sum_j r_j \frac{1}{n} - \beta \frac{1}{n} \sum_{j \notin D} r_j \right) = 1 + \beta \sum_{j \notin D} r_j - \beta \sum_{j \notin D} r_j = 1$$

2 (a) Code: 3_2_a.py

Top 5 nodes

```
  PageRank score         id
[(0.0020202911815182184, 263),
 (0.0019433415714531497, 537),
 (0.0019254478071662631, 965),
 (0.001852634016241731 , 243),
 (0.0018273721700645144, 285)]
```

Bottom 5 nodes

```
  PageRank score          id
[(0.00038779848719291705, 408),
 (0.0003548153864930145 , 424),
 (0.00035314810510596274, 62),
 (0.0003513568937516577 , 93),
 (0.0003286018525215297 , 558)]
```

2 (b) Code: 3_2_b.py

Top 5 nodes with the highest hubbiness score

```
  Hubbiness score     id
[(1.0,                840),
 (0.9499618624906543, 155),
 (0.8986645288972263, 234),
 (0.8634171101843789, 389),
 (0.8632841092495218, 472)]
```

Bottom 5 nodes with the lowest hubbiness score

```
  Hubbiness score       id
[(0.07678413939216452,  889),
 (0.0660265937341849,   539),
 (0.06453117646225177,  141),
 (0.057790593544330145, 835),
 (0.04206685489093652,  23)]
```

Top 5 nodes with the highest authority score

```
  Authority score     id
[(1.0,                893),
 (0.9635572849634398, 16),
 (0.9510158161074015, 799),
 (0.9246703586198443, 146),
 (0.899866197360405,  473)]
```

Bottom 5 nodes with the lowest authority score

```
  Authority score      id
[(0.08571673456144875, 910),
 (0.08171239406816942, 24),
 (0.07544228624641901, 462),
 (0.06653910487622794, 135),
 (0.05608316377607618, 19)]
```

3 (a)

If there is 0 or 1 element in C_i, it is a clique by definition. If there are at least two elements, any pair of elements share a common factor i, so any pair is connected, so C_i is a clique.

3 (b)

Condition: $i \leq 10^6$ and i is prime ⇔ C_i is a maximal clique.

Proof of ⇒: Suppose C_i is not maximal. $\exists j \notin \mathrm{C}_i \ni C_i + \{\mathrm{j}\}$ is a clique. I.e. j is not a multiple of i, but j and i have a common factor > 1. I.e. i is not prime. Contradiction.

Proof of ⇐: Suppose i is not prime or $i > 10^6$. First possibility: $\exists j|$i, $j \neq 1$, $j \notin C_i$.Therefore, $\forall k \in C_i$, $j|k$. So $C_i + \{j\}$ is a clique but $|C_i + \{\mathrm{j}\}| > |C_i|$. Contradiction. Second possibility: C_i is empty so adding any number in G would make a larger clique. Contradiction.

3 (c)

Note: If $\mathrm{G} = \mathbb{N} - \{1\}$, the statement in question is wrong, because $|C_2| = |\mathbb{N}| = |C_3|$.

Lemma: The largest possible clique has 500000 elements. Proof: Because a pair of consecutive integers are always coprime, we can only choose one out of each of {2, 3}, {4, 5}, ..., {999998, 999999}, {1000000} to form a clique. Therefore, we can at most have 500000 elements.

$|C_2| = 500000$ so C_2 is one of the largest cliques.

Uniqueness: By the lemma, the largest cliques include either 2 or 3. If we pick 3, the best we can do is to include all multiples of 3, forming C_3, which has less than 500000 elements.

So C_2 is the unique largest clique.

Note:

A key to understanding this problem is that $\rho(S) = \frac{|E[S]|}{|S|}$ is half of the average degree.

A key to understanding the pseudocode is that $\tilde{S}$ is to keep the best S.

4 (a)

i.

$2|\mathrm{E[S]}| = \sum_{i \in S} deg_S(i)$ (Each edge in $E[S]$ contributes to 2 among all $deg_s(i)$.)

$\geq \sum_{i \in S \setminus A} deg_S(i)$ (S\A is a subset of S.)

$> \sum_{i \in S \setminus A} 2(1+\epsilon)\frac{|E[S]|}{|S|}$ (from the definition of A(S) in the pseudocode)

$= |S \setminus A| 2(1+\epsilon)\frac{|E[S]|}{|S|}$

Therefore,

$\frac{1}{1+\epsilon}|\mathrm{S}| > |\mathrm{S}\setminus\mathrm{A}| = |S| - |A|$ (A is a subset of S.)

$\Rightarrow |\mathrm{A}| > \frac{\epsilon}{1+\epsilon}|\mathrm{S}|$

ii.

Since $\frac{1}{1+\epsilon}|S| > |S \setminus A|$, after each iteration, $|S|$ decreases by at least $1+\epsilon$ times, so there are $\mathrm{O}(\log_{1+\epsilon} n)$ iterations.

4 (b)

i.

Say $\deg_{S*}(v) < \rho^*(G)$. Then $\frac{deg_{S*}(v)}{1} < \frac{|E[S^*]|}{|S^*|}$. Then $\rho(S^* \setminus v) = \frac{|E[S^*]| - deg_{S^*}(v)}{|S^*| - 1} > \rho(S^*)$, contradiction.

ii.

$2(1+\epsilon)\rho(S) \geq deg_S(v)$ ($v \in \mathrm{A}$)

$\geq \rho^*(G)$ (by i., since $v \in S^*$)

iii.

$\rho^*(\tilde{S}) \geq \rho(S)$ (because of the algorithm)

$\geq \frac{1}{2(1+\epsilon)}\rho^*(G)$ (by ii.)