Problem 1

a) *15 points, 5 points for each part of similarity.*
for each part, [-3] for incorrect probability; [-1] for incorrect expected number.
If it is 50% similar,
Probability that a pair is a candidate: $1 - (1 - 0.5^2)^4 =$ **0.6836**
Expected number of candidate pairs: 0.6936 * 1000 = **684**
If it is 20% similar,
Probability that a pair is a candidate: $1 - (1 - 0.2^2)^4 =$ **0.1507**
Expected number of candidate pairs: 0.1507 * 1000000 = **150653**
If it is 0% similar,
Probability that a pair is a candidate: $1 - (1 - 0^2)^4 =$ **0**
Expected number of candidate pairs: 0 * $({}_{1000000}C_2 - 1000 - 1000000) =$ **0**

b) *15 points, 5 points for each part of similarity.*
for each part, [-3] for incorrect probability; [-1] for incorrect expected number.
If it is 50% similar,
Probability that a pair is a candidate: $1 - (1 - 0.5^4)^2 =$ **0.1211**
Expected number of candidate pairs: 0.1211 * 1000 = **121**
If it is 20% similar,
Probability that a pair is a candidate: $1 - (1 - 0.2^4)^2 =$ **0.003197**
Expected number of candidate pairs: 0.003197 * 1000000 = **3197**
If it is 0% similar,
Probability that a pair is a candidate: $1 - (1 - 0^4)^2 =$ **0**
Expected number of candidate pairs: 0 * $({}_{1000000}C_2 - 1000 - 1000000) =$ **0**

c) *5 points*
[-3] for an answer without any justification.
Both (a) and (b) can be a correct answer.
If you choose (a), one reasonable justification can be that (a) gives more 50% similar pairs (684 vs 121).
If you choose (b), one reasonable justification can be that (b) gives a higher proportion of 50% similar pairs (684/(684 + 150653) = 0.45% vs 121/(121 + 3197) = 3.6%).

Problem 2

[-3] for each incorrect grouping.

The 7 groupings are as follows: ({cluster}, centroid)

Step	2	3	5	7	11	13	17	19	23	29
1	({2, 3}, 4)		5	7	11	13	17	19	23	29
2	({2, 3}, 4)		({5, 7}, 6)		11	13	17	19	23	29
3	({2, 3}, 4)		({5, 7}, 6)		({11, 13}, 12)		17	19	23	29
4	({2, 3}, 4)		({5, 7}, 6)		({11, 13}, 12)		({17, 19}, 18)		23	29
5	({2, 3, 5, 7}, 4.25)				({11, 13}, 12)		({17, 19}, 18)		23	29
6	({2, 3, 5, 7}, 4.25)				({11, 13}, 12)		({17, 19, 23}, 19.67)			29
7	**({2, 3, 5, 7}, 4.25)**				**({11, 13, 17, 19, 23}, 16.6)**					**29**

Problem 3

a) 15 points, 1.5 points for each cosine.
[-1.5] for each incorrect cosine, [-10] for using an incorrect equation.
The cosines between different vectors are as follows:

	A	B	C	D	E
A	1	$2/(5^{0.5}4^{0.5}) = 1/5^{0.5}$ = **0.45**	$3/(5^{0.5}5^{0.5}) = 3/5$ = **0.6**	$2/(5^{0.5}4^{0.5}) = 1/5^{0.5}$ = **0.45**	$3/(5^{0.5}5^{0.5}) = 3/5$ = **0.6**
B	$2/(5^{0.5}4^{0.5}) = 1/5^{0.5}$ = **0.45**	1	$1/(4^{0.5}5^{0.5}) = 1/(2\ 5^{0.5})$ = **0.22**	$2/(4^{0.5}4^{0.5}) = 1/2$ = **0.5**	$3/(4^{0.5}5^{0.5}) = 3/(2\ 5^{0.5})$ = **0.67**
C	$3/(5^{0.5}5^{0.5}) = 3/5$ = **0.6**	$1/(4^{0.5}5^{0.5}) = 1/(2\ 5^{0.5})$ = **0.22**	1	$2/(5^{0.5}4^{0.5}) = 1/5^{0.5}$ = **0.45**	$1/(5^{0.5}5^{0.5}) = 1/5$ = **0.2**
D	$2/(5^{0.5}4^{0.5}) = 1/5^{0.5}$ = **0.45**	$2/(4^{0.5}4^{0.5}) = 1/2$ = **0.5**	$2/(5^{0.5}4^{0.5}) = 1/5^{0.5}$ = **0.45**	1	$2/(5^{0.5}4^{0.5}) = 1/5^{0.5}$ = **0.45**
E	$3/(5^{0.5}5^{0.5}) = 3/5$ = **0.6**	$3/(4^{0.5}5^{0.5}) = 3/(2\ 5^{0.5})$ = **0.67**	$1/(5^{0.5}5^{0.5}) = 1/5$ = **0.2**	$2/(5^{0.5}4^{0.5}) = 1/5^{0.5}$ = **0.45**	1

b) 10 points, 5 points for each cluster.
[-3] for each incorrect cluster with explanation; [-5] without any explanation.
{A, B, E} and **{A, C, E}**.

c) 15 points, 1.5 points for each Jaccard distance.
[-1.5] for each incorrect Jaccard distance, [-5] for giving Jaccard measures instead.
The cosines between different vectors are as follows:

	A	B	C	D	E
A	0	1 – 2/7 = 5/7 = **0.71**	1 – 3/7 = 4/7 = **0.57**	1 – 2/7 = 5/7 = **0.71**	1 – 3/7 = 4/7 = **0.57**
B	1 – 2/7 = 5/7 = **0.71**	0	1 – 1/8 = 7/8 = **0.88**	1 – 2/6 = 2/3 = **0.67**	1 – 3/6 = 3/6 = **0.5**
C	1 – 3/7 = 4/7 = **0.57**	1 – 1/8 = 7/8 = **0.88**	0	1 – 2/7 = 5/7 = **0.71**	1 – 1/9 = 8/9 = **0.89**
D	1 – 2/7 = 5/7 = **0.71**	1 – 2/6 = 2/3 = **0.67**	1 – 2/7 = 5/7 = **0.71**	0	1 – 2/7 = 5/7 = **0.71**
E	1 – 3/7 = 4/7 = **0.57**	1 – 3/6 = 3/6 = **0.5**	1 – 1/9 = 8/9 = **0.89**	1 – 2/7 = 5/7 = **0.71**	0