CSC 555 - PROGRAMMING ASSIGNMENT P1

PART – 2

Task 1

Hypothesis 1: Characters can be characterized as on the light side or dark side by comparing their homophily to the subnetwork they belong to.

Proposed metrics: The metric proposed to evaluate this hypothesis is **cohesion**. Cohesion measures the degree to which the nodes are directly connected to each other. In order to define the homophily of the structure, the way the main character interacts with all the other characters is important. This should take into account only the direct interactions between the characters. Once this is known, the number of interactions is measured with respect to the average of all the values. This is the threshold. Based on this value, the characters are classified as light or dark. The results for this method are discussed in the section Task 2.

Hypothesis 2: The character of a series, developed to the main character will have the highest degree centrality, however the most integral character of a series will have the highest betweenness.

Proposed metrics: The metrics proposed to evaluate the hypothesis are **degree centrality** and **betweenness**. Degree centrality measures the total number of links that go from the node to another node and the links that go from another node to the given node. These two represent the indegree and the outdegree of a node. The given graph is an undirected graph. Hence the measure of degree centrality is the number of links attached to the node. This measure tells the viewer how important the node is compared to all the other nodes in the network. Betweenness measures how important the node is by removing the node and comparing the path lengths of its neighbors to other nodes before and after removal. If the path lengths are the same, then the node is not important, else, the node is important. The metrics are compared with each other for the top 5 characters. The results are discussed in the section Task 2.

Task 2

Hypothesis 1: To prove hypothesis 1, it is assumed that 'R2-D2' is a light character. This assumption is based on the fact that R2-D2 occurs throughout the prequel and the original series and hence it would be more appropriate to consider it in order to classify characters throughout the series. Using the number of interactions between all the characters with R2-D2, a threshold was set. This threshold is the cohesion of R2-D2 with other characters. The threshold was calculated as the **sum of all interactions of characters with R2-D2 divided by the total number of characters with whom R2-D2 interacted with.** The results obtained are as follows:

Hypothesis 1		
Character	**Actual**	**Predicted**
Chewbacca	Light	Light
Obi Wan	Light	Light
Padme	Light	Light
Qui Gon	Light	Light
Anakin	Light	Light
Emperor	Dark	Dark
Jabba	Light	Dark
Yoda	Light	Dark
Luke	Light	Light
Leia	Light	Light
Darth Vader	Dark	Dark
Han	Light	Light

The above table shows the predictions for the top characters. The threshold used predicts **10 out of 12** characters correctly and hence gives an **accuracy of 83.3%**. Since R2-D2 is not a very important character in the series, few characters are predicted wrongly. The accuracy can be improved by taking into account more than one main character and trying to find patterns in the interactions among the characters.

The same procedure was performed for all the episodes the results obtained are shown below. The problem with this metric is that if the light character and the dark character spoke to each other during fights or argument or if the dark character was a spy and interacted with the light character frequently, it would classify the character as light.

```
[
    {'QUI-GON': 'light', 'JAR JAR': 'light', 'BOSS NASS': 'dark', 'SIO BIBBLE': 'dark', 'ANAKIN': 'light',
    'SHMI': 'dark', 'YODA': 'dark', 'MACE WINDU': 'dark', 'KI-ADI-MUNDI': 'dark', 'PADME': 'dark'},

    {'R2-D2': 'light', 'JAR JAR': 'light', 'OBI-WAN': 'light', 'CAPTAIN TYPHO': 'dark', 'PADME': 'light',
    'JOBAL': 'dark', 'C-3PO': 'light', 'OWEN': 'light', 'NUTE GUNRAY': 'dark', 'COUNT DOOKU': 'dark'},

    {'R2-D2': 'light', 'ANAKIN': 'light', 'ODD BALL': 'dark', 'EMPEROR': 'dark', 'COUNT DOOKU': 'dark', 'GENERAL GRIEVOUS': 'dark',
    'BAIL ORGANA': 'light', 'C-3PO': 'dark', 'MACE WINDU': 'dark', 'YODA': 'light', 'PADME': 'light', 'TION MEDON': 'dark'},

    {'R2-D2': 'light', 'CHEWBACCA': 'light', 'BERU': 'dark', 'OWEN': 'dark', 'C-3PO': 'light',
    'LEIA': 'light', 'OBI-WAN': 'light', 'HAN': 'light', 'RED LEADER': 'dark', 'RED TEN': 'dark'},

    {'R2-D2': 'light', 'CHEWBACCA': 'light', 'HAN': 'light', 'OBI-WAN': 'light', 'C-3PO': 'dark',
    'DACK': 'light', 'WEDGE': 'light', 'ZEV': 'dark', 'YODA': 'light', 'LANDO': 'light'},

    {'R2-D2': 'light', 'CHEWBACCA': 'light', 'BOUSHH': 'dark', 'LEIA': 'light', 'C-3PO': 'light',
    'LUKE': 'light', 'LANDO': 'dark', 'MON MOTHMA': 'dark', 'ADMIRAL ACKBAR': 'dark'},

    {'CHEWBACCA': 'light', 'BB-8': 'light', 'CAPTAIN PHASMA': 'dark', 'POE': 'light', 'REY': 'light',
    'HAN': 'light', 'BALA-TIK': 'dark', 'MAZ': 'dark', 'ADMIRAL ACKBAR': 'dark', 'KYLO REN': 'dark'}
]
```

The results were obtained by using different characters as the central main character for each episode since all the characters were not present in all the episodes. The main character was chosen based on the number of scenes the character appeared in in the episode.

The above results can be used to conclude that it is not possible to determine all the characters correctly but most of the characters can be determined correctly and this validates the hypothesis that characters can be categorized as light or dark by comparing the homophily to the subnetwork they belong to.

Hypothesis 2: To validate the hypothesis, the degree centrality and the betweenness are calculated for all the characters and checked whether the top 5 characters in both of them match each other. This is then compared to the overall network. The results are shown below:

Episode 1				Episode 2			
Character	Betweenness	Character	Degree Centrality	Character	Betweenness	Character	Degree Centrality
QUI-GON	0.322613523	QUI-GON	26	ANAKIN	0.383382209	ANAKIN	21
ANAKIN	0.169883628	ANAKIN	23	OBI-WAN	0.314801093	OBI-WAN	18
JAR JAR	0.14583615	JAR JAR	18	PADME	0.178946725	PADME	17
NUTE GUNRAY	0.125431384	PADME	18	KI-ADI-MUNDI	0.0625	MACE WINDU	10
PADME	0.103421387	OBI-WAN	13	JAR JAR	0.062130376	YODA	10

Episode 3				Episode 4			
Character	Betweenness	Character	Degree Centrality	Character	Betweenness	Character	Degree Centrality
OBI-WAN	0.18631815	ANAKIN	14	LUKE	0.32239229	LUKE	15
ANAKIN	0.184748102	OBI-WAN	13	LEIA	0.230498866	LEIA	12
EMPEROR	0.18323844	BAIL ORGANA	12	HAN	0.176190476	C-3PO	10
BAIL ORAGANA	0.181875431	EMPEROR	11	BIGGS	0.041666667	R2-D2	9
YODA	0.115924776	PADME	10	C-3PO	0.040646259	BIGGS	8

Episode 5				Episode 6			
Character	Betweenness	Character	Degree Centrality	Character	Betweenness	Character	Degree Centrality
LUKE	0.416217817	LUKE	12	LUKE	0.596881092	LUKE	15
DARTH VADER	0.324675325	HAN	11	DARTH VADER	0.207602339	C-3PO	11
HAN	0.111758943	DARTH VADER	11	C-3PO	0.064230019	HAN	9
WEDGE	0.1	LEIA	10	LANDO	0.058089669	LEIA	9
LEIA	0.073989519	C-3PO	10	AD. ACKBAR	0.043274854	LANDO	9

Episode 7				Overall			
Character	Betweenness	Character	Degree Centrality	Character	Betweenness	Character	Degree Centrality
POE	0.299110334	POE	16	OBI-WAN	0.205276165	ANAKIN	42
KYLO REN	0.221244311	FINN	14	C-3PO	0.168103946	OBI-WAN	37
REY	0.119215118	HAN	14	ANAKIN	0.158361913	C-3PO	36
BB-8	0.089700855	CHEWBACCA	12	LUKE	0.141683162	PADME	34
FINN	0.082413142	BB-8	12	HAN	0.096396123	QUI-GON	27

From the above results, it can be seen that the most central character is often not the most integral character. Checking the results for the overall episodes table, it is seen that the **Anakin is the main character with degree centrality = 42** but the most **integral character is Obi Wan with betweenness = 0.2052.** Hence the hypothesis can be proved that the character of a series, developed to be the "main" character will have the highest degree centrality, however the most "integral" character of a series will have the highest betweenness.

Task 3

Cliquishness and path lengths:

The cliquishness is given by the clustering coefficient which measures the degree to which a node's neighbors are connected to one another. The cliquishness of each episode is shown below for the bottom 5 characters and the top 5 characters.

Episode 1	Character	Cluster Coefficient
	PK-4	0
	QUI-GON	0.255384615
Bottom 5	ANAKIN	0.280632411
	TC-14	0.3
	NUTE GUNRAY	0.333333333
	JABBA	0.80952381
	SHMI	0.821428571
Top 5	DARTH MAUL	0.833333333
	YODA	0.933333333
	R2-D2	1

Episode 2	Character	Cluster Coefficient
	PLO KOON	0
	ANAKIN	0.219047619
Bottom 5	OBI-WAN	0.300653595
	PADME	0.323529412
	JAR JAR	0.577777778
	JANGO FETT	0.666666667
	C-3PO	0.8
Top 5	SENATOR ASK AAK	0.80952381
	BAIL ORGANA	0.857142857
	R2-D2	1

Episode 3	Character	Cluster Coefficient
	DARTH VADER	0
	GENERAL GRIEVOUS	0.333333333
Bottom 5	ANAKIN	0.351648352
	EMPEROR	0.363636364
	OBI-WAN	0.371794872
	PADME	0.511111111
	C-3PO	0.638888889
Top 5	R2-D2	0.75
	MACE WINDU	0.8
	QUI-GON	1

Episode 4	Character	Cluster Coefficient
	GREEDO	0
	LUKE	0.342857143
Bottom 5	LEIA	0.424242424
	HAN	0.535714286
	RED LEADER	0.571428571
	DARTH VADER	0.666666667
	OBI-WAN	0.761904762
Top 5	GOLD LEADER	0.8
	BERU	0.9
	CHEWBACCA	1

Episode 5		
	Character	Cluster Coefficient
Bottom 5	EMPEROR	0
	LUKE	0.348484848
	DARTH VADER	0.418181818
	PIETT	0.5
	HAN	0.545454545
Top 5	OBI-WAN	0.666666667
	BOBA FETT	0.761904762
	CHEWBACCA	0.777777778
	LANDO	0.892857143
	R2-D2	1

Episode 6		
	Character	Cluster Coefficient
Bottom 5	PIETT	0
	DARTH VADER	0.2
	LUKE	0.304761905
	C-3PO	0.6
	BOUSHH	0.666666667
Top 5	ADMIRAL ACKBAR	0.761904762
	LEIA	0.777777778
	R2-D2	0.821428571
	MON MOTHMA	0.928571429
	CHEWBACCA	1

Episode 7		
	Character	Cluster Coefficient
Bottom 5	LUKE	0
	GENERAL HUX	0.3
	POE	0.391666667
	KYLO REN	0.4
	BB-8	0.46969697
Top 5	JESS	0.7
	C-3PO	0.711111111
	LEIA	0.755555556
	SNAP	0.785714286
	R2-D2	1

Overall network		
	Character	Clustering Coefficient
Bottom 5	PLO KOON	0
	ANAKIN	0.195121951
	OBI-WAN	0.217717718
	PADME	0.242424242
	LUKE	0.245014245
Top 5	BOUSHH	0.833333333
	OWEN	0.857142857
	CAPTAIN TYPHO	0.866666667
	SIO BIBBLE	0.928571429
	PK-4	1

The average path lengths of all the social networks is shown below:

Path lengths	
Episode	Path Length
1	2.105263158
2	2.138257576
3	1.014492754
4	0.957142857
5	2.061904762
6	1.947368421
7	2.082621083
Overall	1.332022932

Randomizing the graph

The social network graph for all the episodes are randomized at a rewire probability. When the graphs are randomized, the **average clustering coefficient** of the graph **decreases**. This means that as the probability increases, the graph tends to stop behaving as a small world network.

The original graph and the randomized graph are shown below (rewire probability = 0.1):
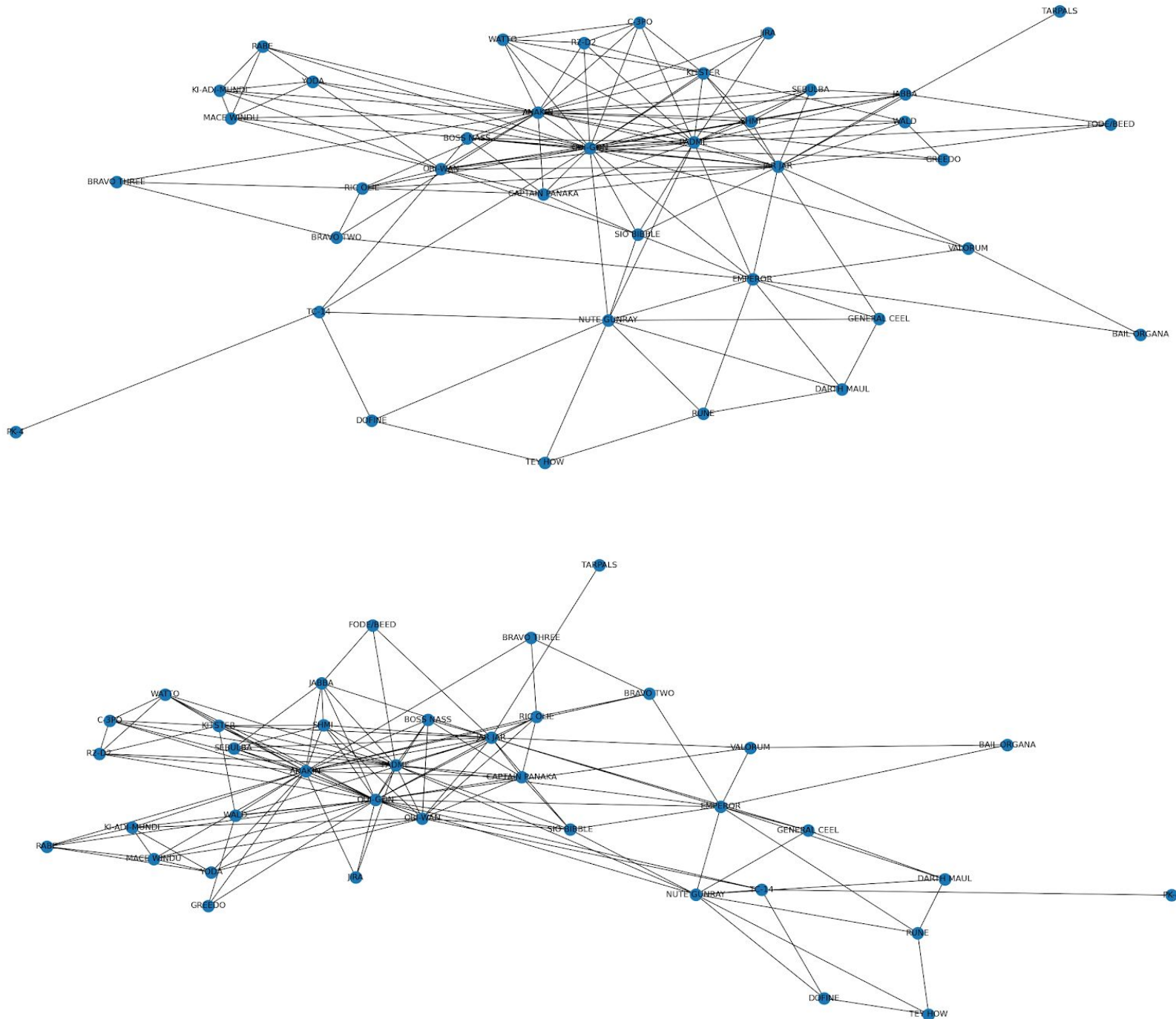
Here it can be seen that Tarpals was connected to Jabba in the original graph and after randomizing, Tarpals is connected to Jar Jar.
The following table shows the results for a rewire probability of 0.1:

Clustering of the graph at p=0.1		
Episode	Original	Randomized
1	0.711574591	0.649972754
2	0.677909346	0.663339251
3	0.599563992	0.575995338
4	0.663872491	0.644808474
5	0.628581736	0.550814265
6	0.609444444	0.571447441
7	0.701772405	0.673068701
Overall	0.686802542	0.659222338

Examining the results for different rewiring probabilities, it was seen that the **difference in the cliquishness** of the two graphs grew **larger and larger**. At probability 0.6, the following results are obtained:

Clustering of the graph at p=0.6		
Episode	Original	Randomized
1	0.711574591	0.595481748
2	0.677909346	0.601383882
3	0.599563992	0.472432234
4	0.663872491	0.616889676
5	0.628581736	0.527475435
6	0.609444444	0.497591575
7	0.701772405	0.576142993
Overall	0.686802542	0.578688244

At probabilities > 0.5, the difference in the cliquishness tends to increase significantly and the graph tends to stop being a small world network.

Task 4

The weak ties are chosen by using both the **number of interactions and the number of scenes the character appeared in the episode.** The two values are summed up and taken as the metric. For each episode, these values for each character are calculated and sorted to find the last three values. The same procedure is performed for the overall episodes. The results found are shown below:

Task 4		
Episode	Weak ties	Count
1	PK-4	4
	GREEDO	6
	VALORUM	8
2	PK-4	4
	PLO KOON	4
	SOLA	8
3	PLO KOON	3
	JAR JAR	5
	DARTH VADER	6
4	GREEDO	4
	CAMIE	6
	GOLD FIVE	9
5	EMPEROR	4
	DERLIN	6
	OBI-WAN	7
6	YODA	4
	OBI WAN	4
	ANAKIN	5
7	LUKE	3
	COLONEL DATOO	5
	LIEUTENANT MITAKA	6

After examining the results shown above, it can be seen that some of the characters like Plo Koon, Greedo and Obi Wan occur as weak ties in more than one episode. But in the analysis from Task 2, it can be seen that Obi-Wan is one of the main characters in the series. These weak ties may have occurred due to the fact that Obi-Wan dies in the series. Some characters appear only in certain episodes and fade away as the series progresses. Examining just the individual episodes in this case will not help us determine the weak ties throughout the series. Hence, considering the overall episodes and interactions for all the characters, the result in the following result is obtained.

Episode	Weak ties	Count
Overall	PLO KOON	5
	PK-4	6
	SOLA	8

On examining the whole network, it can be seen that the top weak ties are Plo Koon, PK-4 and Sola. Rankings for all episodes:

Episode	Order (weakest < weaker < weak)
Episode 1	PK-4 < Greedo < Valorum
Episode 2	PK-4 = Plo Koon < Sola
Episode 3	Plo Koon < Jar Jar < Darth Vader
Episode 4	Greedo < Camie < Gold Five
Episode 5	Emperor < Derlin < Obi Wan
Episode 6	Yoda = Obi Wan < Anakin
Episode 7	Luke < Colonel Datoo < Lt. Mitaka
Overall	**Plo Koon < PK-4 < Sola**

Task 5

Based on the data given, the following results were obtained for each episode.
X-axis: Character names
Y-axis: number of interactions + number of scenes the character appeared in

Episode 1:
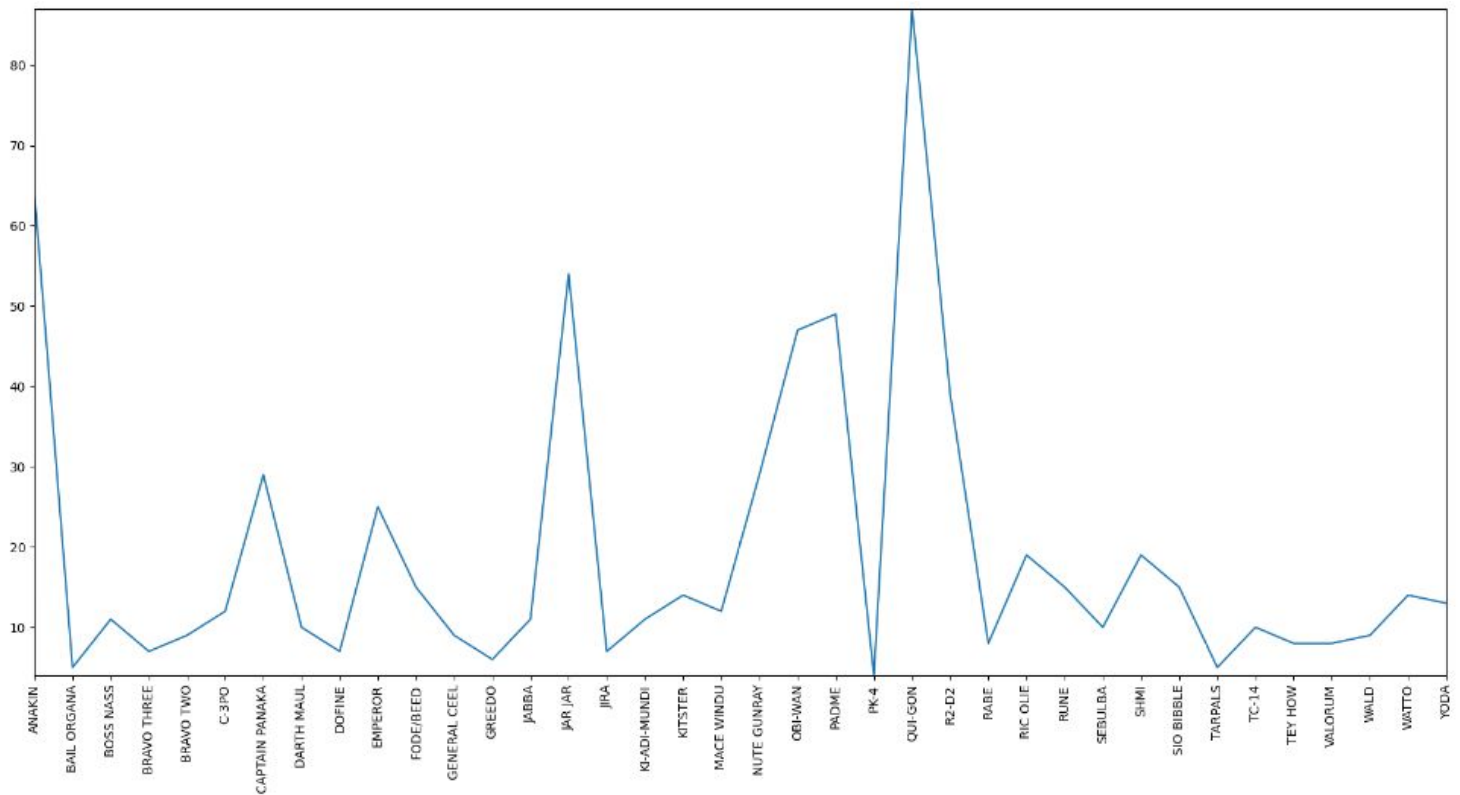
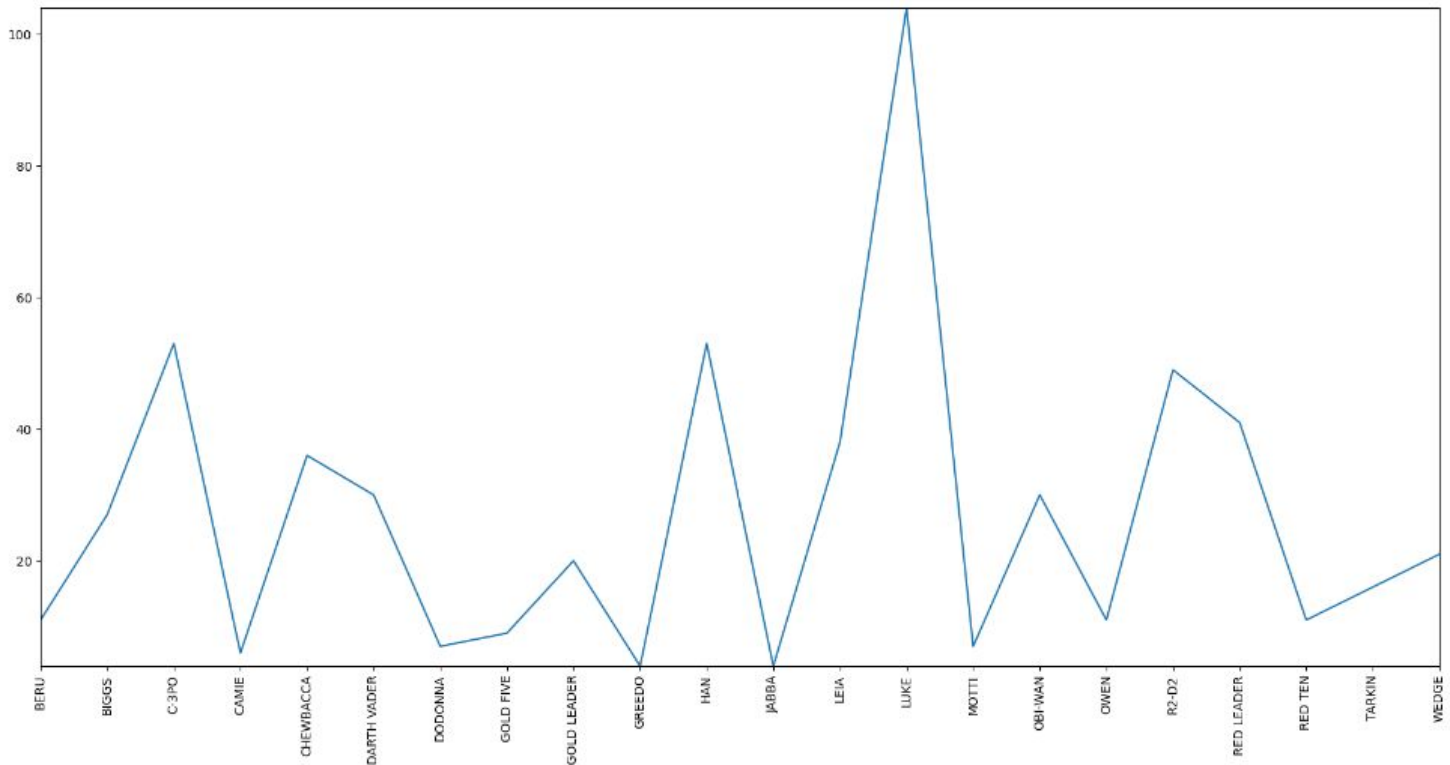The main characters are in the order: Qui Gon > Anakin > Jar Jar > Obi Wan > Padme

Episode 2:
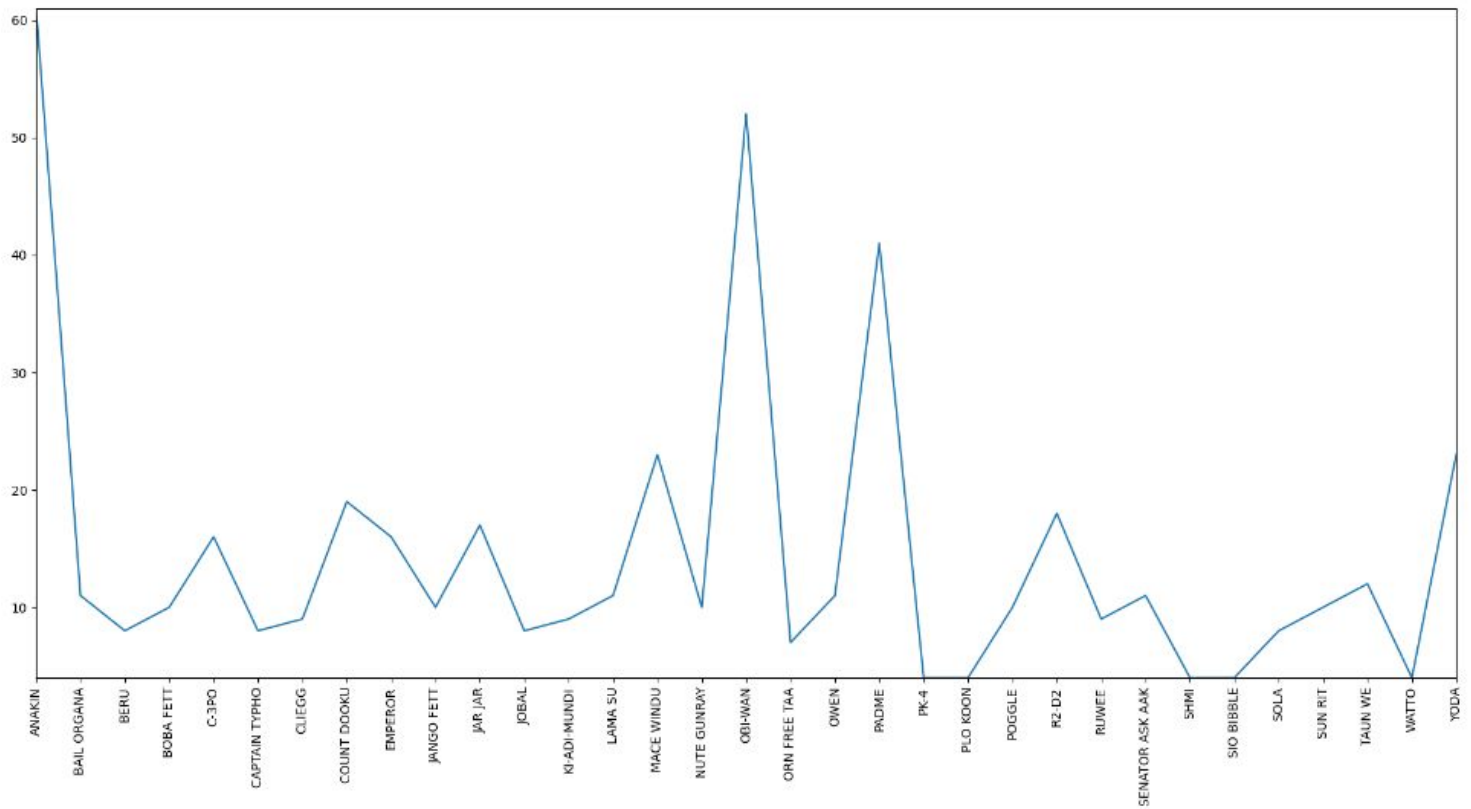
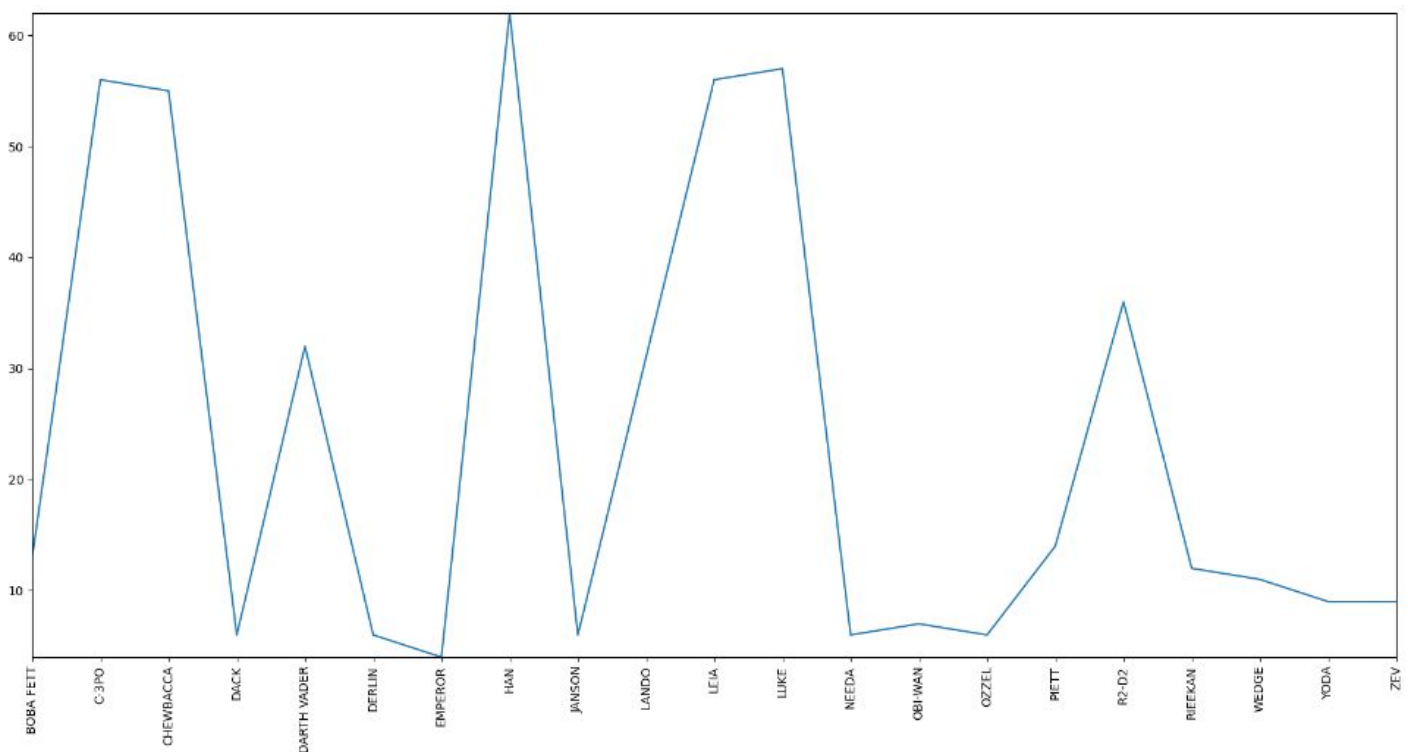The main characters in this are in the order: Anakin > Obi Wan > Padme

Episode 3:
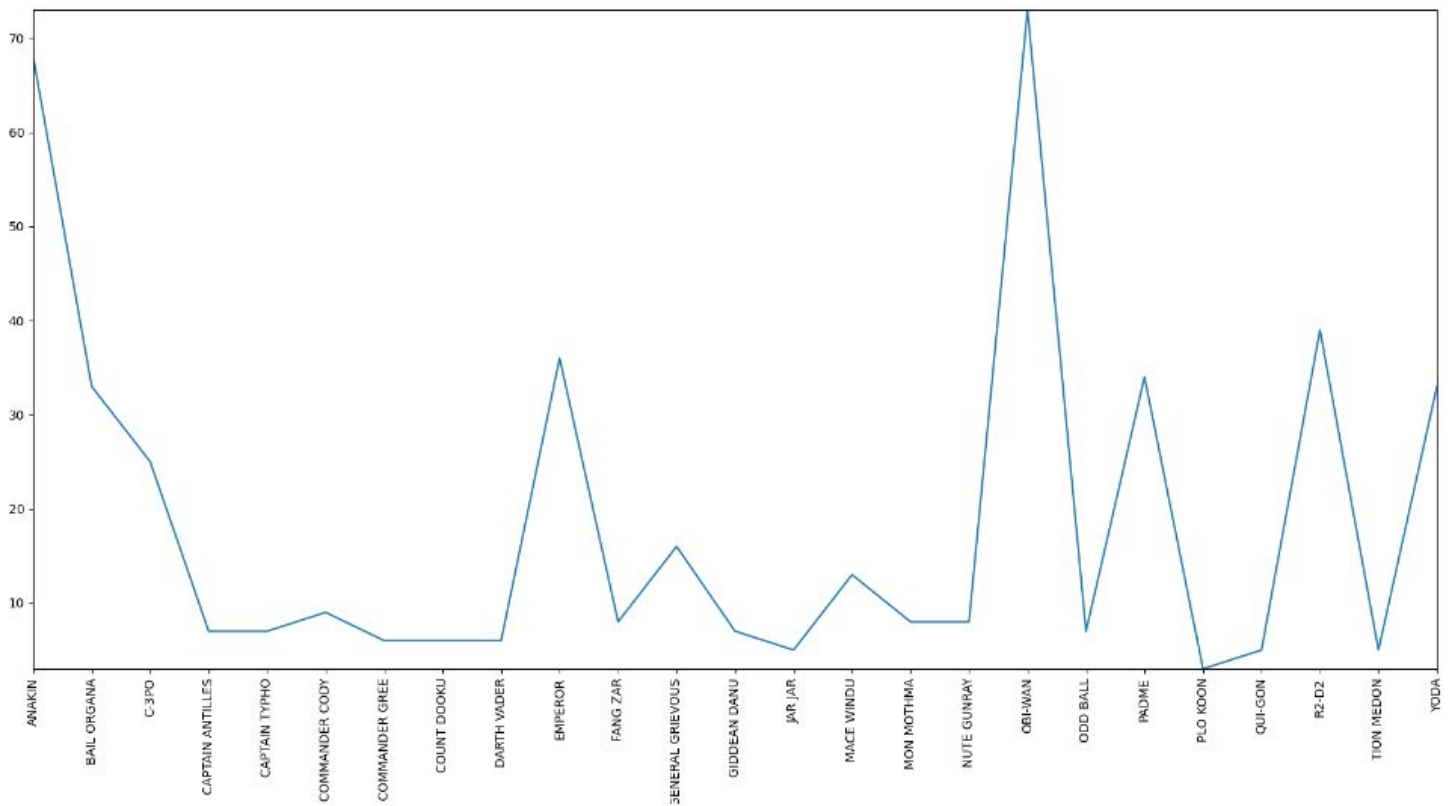
The main characters in this episode are: Obi Wan > Anakin > R2-D2 > Emperor > Padme

Episode 4:

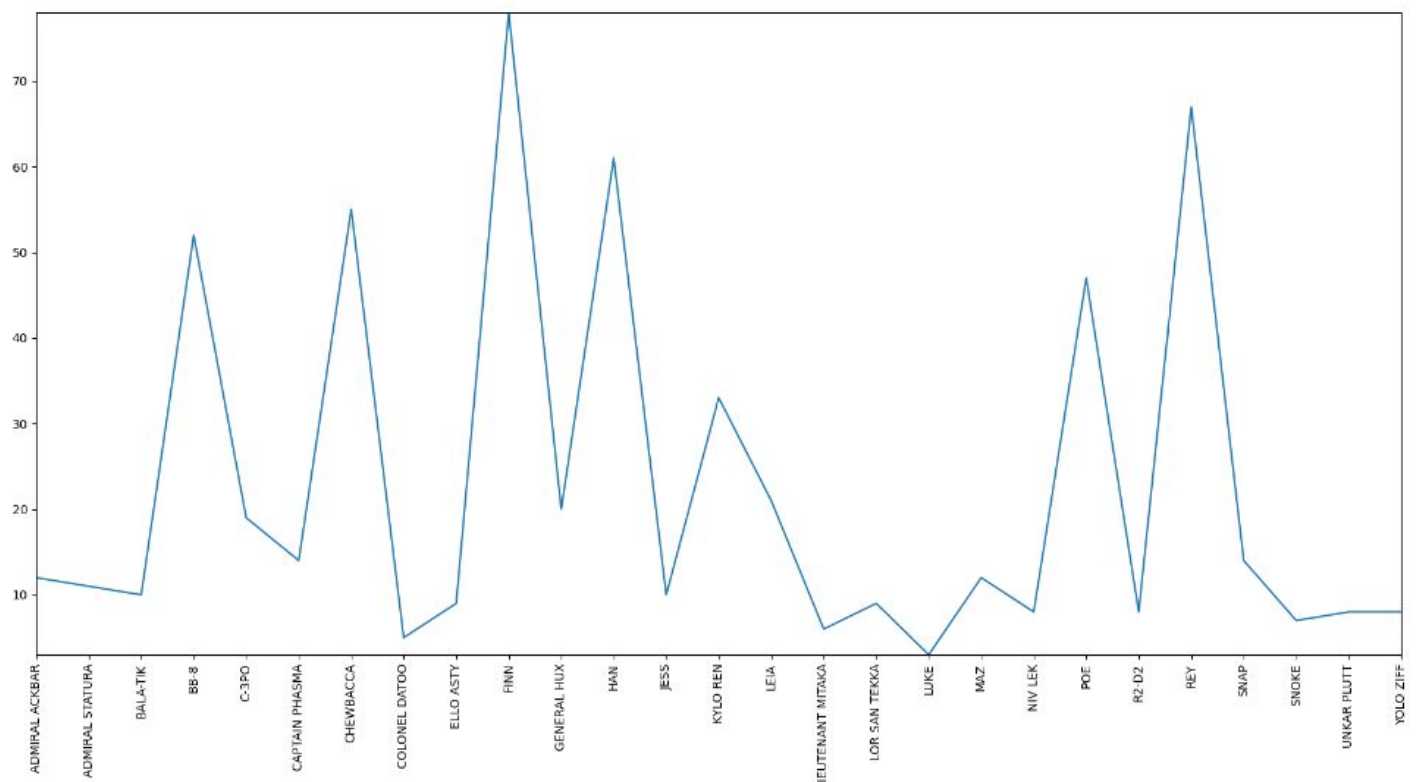The main characters are in the order: Luke > Leia > C-3PO > Han > R2-D2

Episode 5:

The main characters are in the order: Han > C-3PO = Luke > Chewbacca =Leia
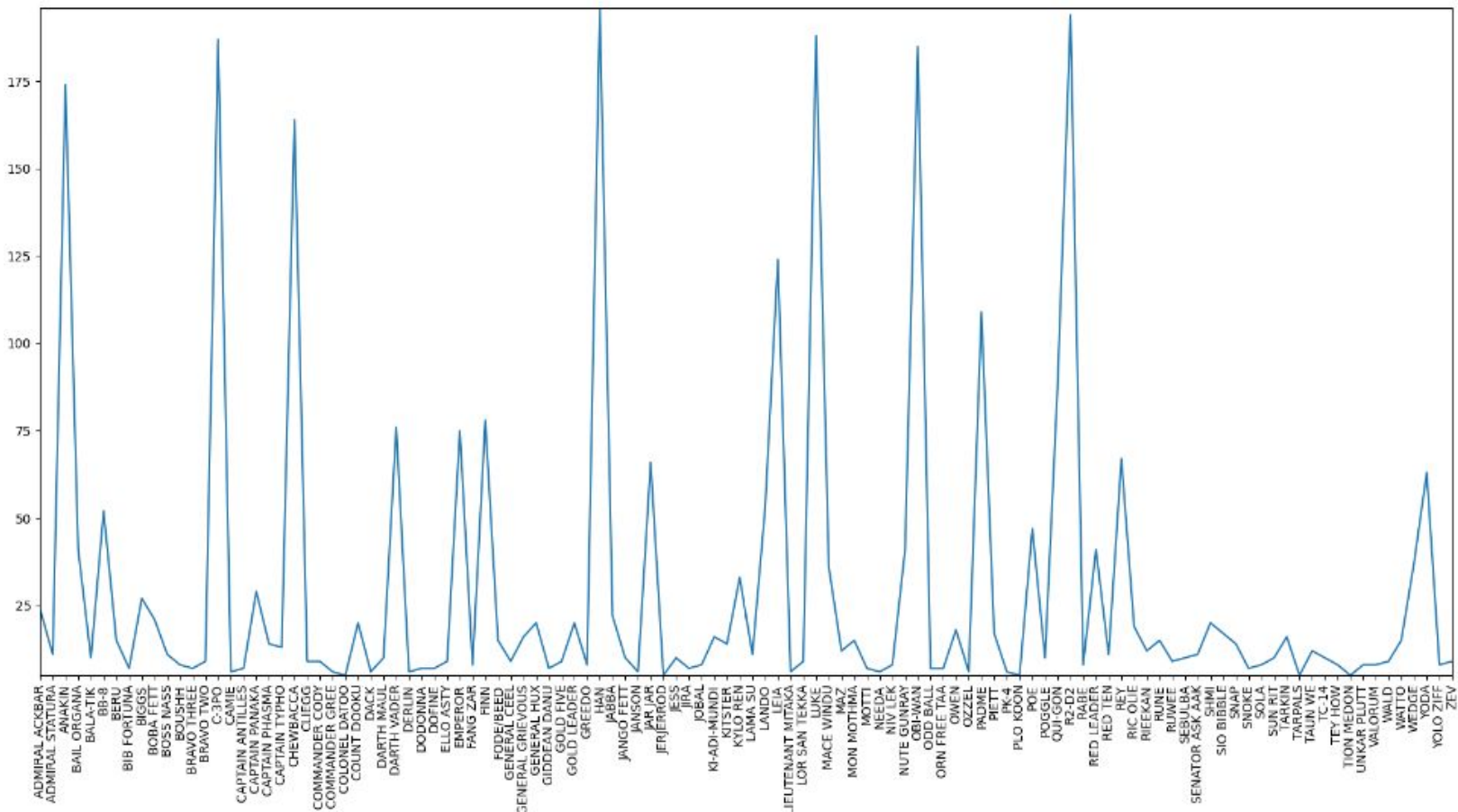
Episode 6:
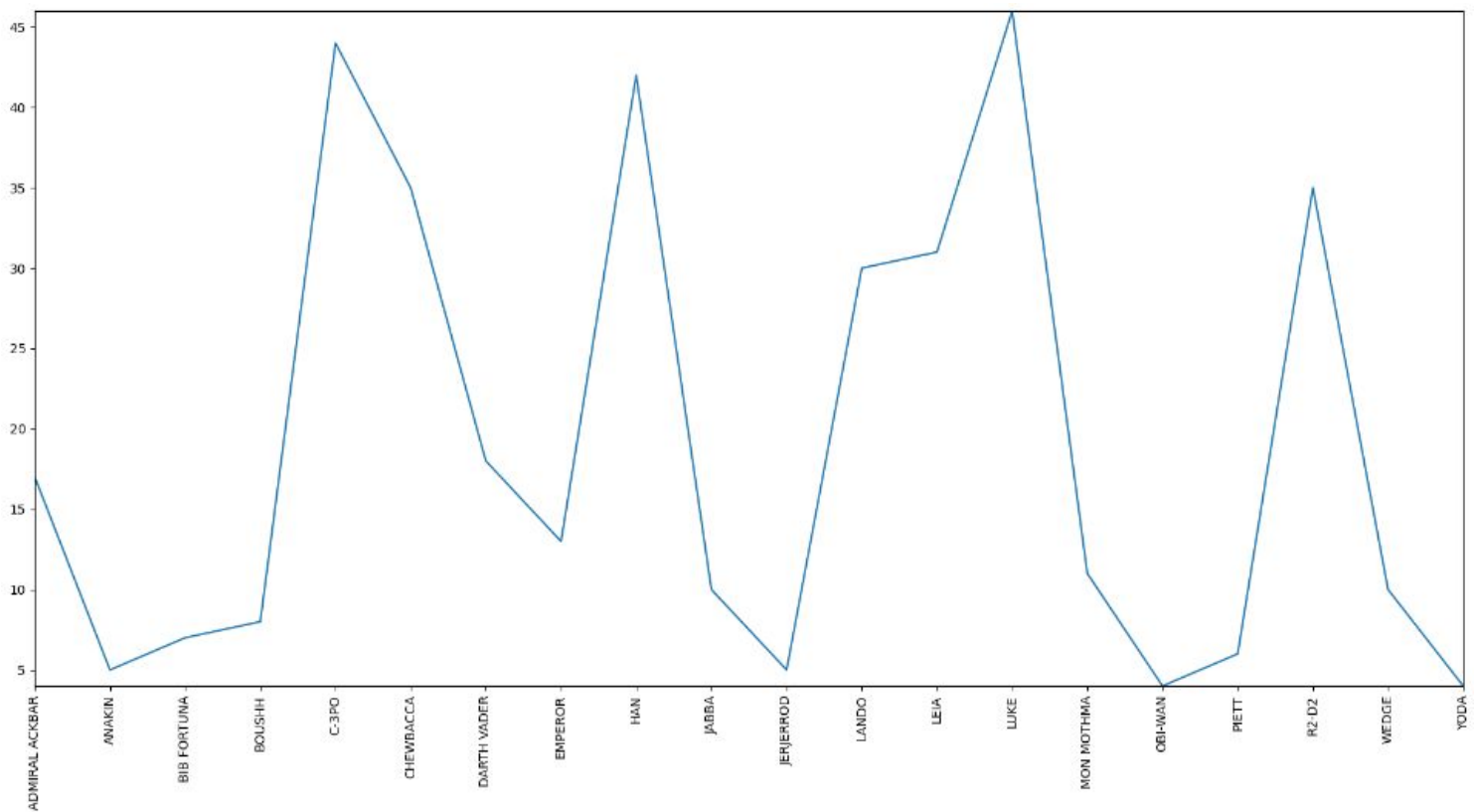
The main characters are in the order: Luke > C-3PO > Han > R2-D2 > Leia

Episode 7:

The main characters are in the order: Finn > Rey > Han > Chewbacca > BB-8 > POE

Overall network:

In the overall network, it can be seen that the main characters are:
Han > R2-D2 > C-3PO = Luke > Obi Wan > Anakin > Chewbacca > Leia > Piett

Conclusion:
1. A few of the characters who appear as important in the overall network do not appear as important in the individual episodes. This tells us that the way important characters are determined is relative to all the characters in the same episode. Hence, the characters who may seem central in the particular episode may not be so.
2. Anakin was more integral in the prequel (before he became Darth Vader) and it seems like Obi Wan and Luke took over the series once Darth Vader was born.
3. As the series progresses, one character fades away and another takes over and the story adapts to the new character and the main subnetwork keeps changing accordingly.