

Preparing dataframe for topic visualisation

In [2]:

```
import pandas as pd
```

In [2]:

```
df = pd.read_csv('topic_model.csv')
```

In [3]:

```
df
```

Out[3]:

	Unnamed: 0	Topic	Word	Proportion
0	0	0	danmark	0.934967
1	1	0	forrest	0.034483
2	2	0	førertrøje	0.018754
3	3	0	her	0.011797
4	4	1	mange	0.270477
...
903	903	94	soja	0.038629
904	904	94	import	0.035249
905	905	94	dyrke	0.030903
906	906	94	svin	0.027523
907	907	94	halvere	0.024626

908 rows × 4 columns

In [4]:

```
#giving topics titles
df['Topic'] = df['Topic'].apply(lambda x: 'Denmark as frontrunner' if x == 0 else x)
df['Topic'] = df['Topic'].apply(lambda x: 'Agriculture' if x == 4 else x)
df['Topic'] = df['Topic'].apply(lambda x: 'Green Energy' if x == 24 else x)
df['Topic'] = df['Topic'].apply(lambda x: 'Research and innovation' if x == 29 else x)
df['Topic'] = df['Topic'].apply(lambda x: 'Nature conservation' if x == 30 else x)
df['Topic'] = df['Topic'].apply(lambda x: 'Economy' if x == 33 else x)
df['Topic'] = df['Topic'].apply(lambda x: 'Climate law' if x == 35 else x)
df['Topic'] = df['Topic'].apply(lambda x: 'Transportation' if x == 42 else x)
df['Topic'] = df['Topic'].apply(lambda x: 'Vehicles' if x == 44 else x)
df['Topic'] = df['Topic'].apply(lambda x: 'Investment' if x == 48 else x)
df['Topic'] = df['Topic'].apply(lambda x: 'Reaching the climate goal' if x == 56 else x)
df['Topic'] = df['Topic'].apply(lambda x: 'Climate crisis' if x == 57 else x)
df['Topic'] = df['Topic'].apply(lambda x: 'Denmark in the world' if x == 61 else x)
df['Topic'] = df['Topic'].apply(lambda x: 'Reuse/recycle' if x == 67 else x)
df['Topic'] = df['Topic'].apply(lambda x: 'Ocean' if x == 69 else x)
df['Topic'] = df['Topic'].apply(lambda x: 'Emissions' if x == 71 else x)
df['Topic'] = df['Topic'].apply(lambda x: 'Windmills' if x == 79 else x)
df['Topic'] = df['Topic'].apply(lambda x: 'Food' if x == 83 else x)
df['Topic'] = df['Topic'].apply(lambda x: 'Europe' if x == 86 else x)
df['Topic'] = df['Topic'].apply(lambda x: 'Animal welfare' if x == 88 else x)
df['Topic'] = df['Topic'].apply(lambda x: 'Water' if x == 89 else x)
df['Topic'] = df['Topic'].apply(lambda x: 'Reforestation' if x == 90 else x)
df['Topic'] = df['Topic'].apply(lambda x: 'Dirty energy' if x == 93 else x)
```

In [5]:

```
df = df[df['Topic'].apply(lambda x: isinstance(x, (str, bytes))))]
```

In [6]:

```
df.head()
```

Out[6]:

	Unnamed: 0	Topic	Word	Proportion
0	0	Denmark as frontrunner	danmark	0.934967
1	1	Denmark as frontrunner	forrest	0.034483
2	2	Denmark as frontrunner	førertrøje	0.018754
3	3	Denmark as frontrunner	her	0.011797
34	34	Agriculture	landbrug	0.453833

In [7]:

```
# create excel writer object
writer = pd.ExcelWriter('topic_words.xlsx')
# write dataframe to excel
df.to_excel(writer)
# save the excel
writer.save()
#print('DataFrame is written successfully to Excel File.')
```

In [8]:

```
df.head()
```

Out[8]:

	Unnamed: 0	Topic	Word	Proportion
0	0	Denmark as frontrunner	danmark	0.934967
1	1	Denmark as frontrunner	forrest	0.034483
2	2	Denmark as frontrunner	førertrøje	0.018754
3	3	Denmark as frontrunner	her	0.011797
34	34	Agriculture	landbrug	0.453833

Preparing dataframe for documents' topic distribution

In [6]:

```
df = pd.read_csv('topic_modelling_final.csv')
```

In [11]:

```
#dropping columns that are not relevant  
df = df.drop(columns = ['Unnamed: 0', 'index', 'Unnamed: 0.1.1.1'])
```

In [11]:

```
#renaming columns to topic names  
df = df.rename(columns={'0':'Denmark as frontrunner', '4':'Agriculture', '24':'Clean energy', '29':'Research and innovation', '30':'Nature conservation', '33':'Economy', '35':'Climate law', '42':'Transportation', '44':'Vehicles', '48':'Investment', '56':'Reaching the climate goal', '57':'Climate crisis', '61':'Denmark in the world', '67':'Reuse/recycle', '69':'Ocean', '71':'Emissions', '79':'Windmills', '83':'Food', '86':'European community', '88':'Animal welfare', '89':'Water', '90':'Reforestation', '93':'Dirty energy'})
```

In [12]:

df

Out[12]:

	Unnamed: 0	index	Unnamed: 0.1.1.1	created_at	text	Account	lemma_
0	0	0	7155.0	2021-04-28 12:46:09	I Danmark har vi på de fleste parametre verden...	KD_Mikkelsen	danmark i parametre ' landbrug la
1	1	1	7170.0	2021-04-08 08:50:17	Stærk fælles udmelding fra @venstredk og @N...	KD_Mikkelsen	stærk udmeldin omstilling lan
2	2	2	7177.0	2020-11-30 19:56:48	@DanskIndustri offentliggør her til aften år...	KD_Mikkelsen	offentliggør erhvervsklimaa
3	3	3	7209.0	2019-02-21 21:53:53	Blandt andet derfor skal vi tale dansk landbru...	KD_Mikkelsen	tale dansk lar politiker dy
4	4	4	7220.0	2018-10-26 17:59:47	@jesperborup @P1radio Svært emne tænkte jeg um...	KD_Mikkelsen	svær emne umidi udsendelse er
...
26047	26047	26047	NaN	2018-04-01 16:52:42	Man ved det er 1. april må klima ministeren f...	Carl__Valentin	april må minister ' kontrove
26048	26048	26048	NaN	2018-02-21 17:39:17	Jeg genopstiller ved EP-valget i 2019. Jeg er ...	Carl__Valentin	genopstiller e rigtignok stadi
26049	26049	26049	NaN	2018-02-16 09:41:06	Jeg har ikke tjekket, hvorfor Løkkefonden hedd...	Carl__Valentin	tjekke l løkkefond løkkefond v
26050	26050	26050	NaN	2018-02-04 01:45:40	Ægte grøn - ja så arbejder man for mere kollek...	Carl__Valentin	ægte grø arbejds kollektiv tr
26051	26051	26051	NaN	2018-01-03 08:31:17	Sådan opdrager du en vinder:\n(1) Du lærer b...	Carl__Valentin	opdrager vinde barn liv konku

26052 rows × 104 columns

In [105]:

```
#grouping by each account, taking the mean of their tweets' topic scores
df = df.groupby('Party', as_index=False)['Denmark as frontrunner', 'Agriculture', 'Clean energy',
    'Research and innovation', 'Nature conservation', 'Economy',
    'Climate law', 'Transportation', 'Vehicles', 'Investment',
    'Reaching the climate goal', 'Climate crisis',
    'Denmark in the world', 'Reuse/recycle', 'Ocean', 'Emissions',
    'Windmills', 'Food', 'European community', 'Animal welfare',
    'Water', 'Reforestation', 'Dirty energy'].mean()
```

<ipython-input-105-7e8b053295c5>:2: FutureWarning: Indexing with multiple keys (implicitly converted to a tuple of keys) will be deprecated, use a list instead.

```
df = df.groupby('Party', as_index=False)['Denmark as frontrunner', 'Agriculture', 'Clean energy',
```

In [108]:

```
# create excel writer object
writer = pd.ExcelWriter('topic_dist.xlsx')
# write dataframe to excel
df.to_excel(writer)
# save the excel
writer.save()
print('DataFrame is written successfully to Excel File.')
```

DataFrame is written successfully to Excel File.

Validation of data

In [5]:

```
sample_green = df.text.sample(100)
```

In [17]:

```
pd.set_option("display.max_rows", 100, )
pd.options.display.max_colwidth = 1000
```

In [19]:

```
sample_green[2800]
```

Out[19]:

2800 Ny rapport fra EUs miljøagentur er alarmerende læsning. Danmark bør være foregangsland for #dknatur og... <https://t.co/E50lGNkQqn>

11101 RT @a_steenberg: To glade @radikale forhandlere💚👍\r\n💚2,1 mio tons CO2 reduktion\r\n💚 775.000 grønne biler - ingen stigning i konventionelle \r\n💚 f...

15263 RT @ida_ndergaard: Nogle af nationens sejeste naturtyper aka @ejrn_s @NaturThor @runeengelbreth @mortenDD førte os i dag rundt på Mols i de...

17101 RT @politiken: Venstre og fagforeningen NNF er gået sammen i en sjælden alliance for at sikre en bred landbrugsaftale. <https://t.co/zF5f8Fp...>

17636 Kæmpe respekt til jer, der kæmpede for borgerforslaget om klimaloven 🙏🙏 \r\nNu skal vi gøre genstarten af Danmark grøn... <https://t.co/teuwTr6YlN>

1054 RT @venstredk: Vi er så klar! Venstre stiller med det stærkeste hold til Europa-Parlamentsvalget 26.maj. \r\n✓Bæredygtig fremtid: EU forrest i...

6776 @LarsHFF @DTUtweet Gå nu væk med dine fakta. Fakta hører ikke hjemme i miljødebatten

11006 Wow, vores søsterparti i Holland @D66!! Lad os give den ligeså meget gas herhjemme for uddannelse og klima 🇳🇱 <https://t.co/Vexk49Kk6m>

10503 RT @UngDKaid: BØRN & UNGE KRÆVER OGSÅ DANSK KLIMALOV!💚🇩🇰\r\n\r\nSammen med en række ungdomsorganisationer har vi lavet parallel underskriftsindsam...

14488 Tænker ofte, at jeg skulle takke de unge, som @WWFdk og @NaturThor samlede til konference om biodiversitet. Vi fik... <https://t.co/wnRJIX2k05>

2745 RT @AnneValentina: EU's statsledere har forhandlet om nyt klimamål hele natten! EU💚🇪🇺\r\nKlimakrisen holder også mig vågen om natten - grøn om...

11331 @pdalgd @Heunicke Nej, den mørkegrønne.

2678 RT @PiaOlsen: Vedvarende energi er vejen frem, så vi når et fossilt frit samfund i 2050, og al energi grønt i 2030! SF er så klar💚🇩🇰 #dkgre...

12692 En arrogant udtalelse fra Dan Jørgensen omkring CO2 afdrift : Der findes ikke noget sted i verden, der har introducere... <https://t.co/Rgnzt6GcuP>

14850 RT @spisekammeret: Åbent Landbrug er i år også natur- og miljøtiltag. Her er Martin Merrild, formand for Landbrug & Fødevarer ude og se på...

4713 RT @kraghclaus: Vidste du godt, at klimaminister @DanJørgensen har sin egen engelsksprogede #podcast kaldet #PlanetA? - Her taler han...

5428 @helenehagel @marqann @pelledragsted @Enhedslisten Grønne regneregler, afgifter på flybrændstof, lovgivning mod madspild og meget mere.

16913 .@uffeelbaek: En rigtig god dag for den grønne omstilling💚 En markant styrkelse af Klimarådet, et fokus på grønne... <https://t.co/07KQF8usbR>

4236 RT @Spolitik: Mette F.: "Danmark skal være en grøn stormagt igen." #dkpol #dkgreen <https://t.co/kbP6s7zA8P>

11031 RT @a_steenberg: Ufatteligt at man kan lave de næste ti års transportpolitik uden markante initiativer på klima. Hvor bliver den der 1 mio...

864 RT @larsloekke: Endnu et skridt mod et bedre klima og en grønner fremtid: Regeringen afsætter 50 mio. kr. til test af fremtidens supervind...

8000 Spørgsmål 364 til klimamin:\r\n- Vil ministeren redegøre for, hvordan proceduren er, hvis man ønsker at stoppe Baltic... <https://t.co/9lKqqHTgvG>

7820 Messerschmidt: Klimarådets anbefalinger er helt på månen <https://t.co/D5ZoJ3HNHT> via @DitOverblik.dk

9453 RT @Kira_MPH: Den sorte industri sidder tungt på posterne i kommissionens ekspertgrupper, når der skal rådgives i den grønne omstilling!\r\nDe...

15671 RT @Kira_MPH: Aftale om EUs klimalov. EU🌐\r\n\r\nTeksten siger mindst 55%, men vi ender reelt på et 2030-mål på 52,8% i absolutte emissionsredukt...

4979 RT @TRM_dk: Boligminister @KaareDybvad og klima,- energi- forsyningsminister @DanJoergensen er nu klar med regeringens udspil til ny grøn b...

1123 Regeringen glemmer dyrevelfærd og borgerinddragelse ifb. med deres udrulning af naturnationalparker. Det kræver et... <https://t.co/o/zdCL3m0iMG>

14591 @emilnielsen @metteabildgaard Jeg synes bestemt også det er på sin plads at rose @metteabildgaard & KF💖🙏 Fremad for naturen!

1208 RT @Besenbacher2012: Der er brug en vision for en ny cirkulær plasticøkonomi i DK, hvor vi opnår en genanvendelse tæt på 100%. DK kan blive...

1989 Landbruget skal skære markant i drivhusgasserne allerede i 2025! @SFpolitik vil sætte turbo på udtag af klimaskadel... <https://t.co/o/ZDVgjxAiRc>

12371 RT @svanDGSB: I dag er regeringens klimaplan udkommet. Hermed starter brandingkampagnen for at undergrave det faktum at de ikke vil vægte b...

15866 RT @rasmusnordqvist: Ja vi står i en alvorlig situation med Coronakrisen - men den har hverken aflyst klima- eller biodiversitetskriserne d...

7765 Lad os gøre håndværkerfradraget fleksibelt til energirenovering <https://t.co/L7FC9ibhK1> #grønneboliger #sundfornuft #dkpol

10469 RT @JohansenTine: Enig med @SandKjaer - selvfølgelig skal big tech bidrage til mediernes og demokratiets økosystem #dkmedier #dkpol #ophavs...

16929 RT @AlternativeUnge: Klima og miljø skal selvfølgelig skrives ind i grundloven! Godt at topcheferne nu også har indset det💖 #dkpol \r\n\r\n<https://t.co/...>

9415 RT @SFpolitik: Det går ikke at udskyde naturpakken. Det haster med flere naturnationalparker. Og naturen skal ligesom klimaet have en lov 💖...

6210 RT @DanJoergensen: Jeg er glad for, at @klimaraadet mener, @regeringDK har taget væsentlige skridt mod 70%-målet. Men de siger også at vi l...

7632 RT @ChristianHansez: Nok om Paludan, selvom det passer visse politikere fint... \r\n\r\nLad os komme tilbage på sporet og snakke klima. Vi har t...

11826 Så er KulturNatten i gang. 😊 Hos DF en aften med klima og sund fornuft🙌 DK

3795 RT @KlimaMin: Danmark er nu nummer 1 i toneangivende globalt bæredygtighedsindeks, som særligt roser Danmark for at lede kampen mod klimafo...

5117 RT @PlasticChange: Hurra! En tredobling af afgifter på

bæreposer og engangsservice er en stor sejr for miljøet og et vigtigt skridt i at gå...

16584 RT @TorstenGejl: Stor dag for Å. \r\nFor et par år siden var 70% CO2 reduktion i 2030 en grøn drøm i Alternativet. \r\nNu er det en politisk aftale...

10709 RT @byplejersken: Hænger producenterne bevidst i klimahåndbremsen og hvorfor er bæredygtighedscertificering ikke nok til at gøre den sorte...

999 RT @venstredk: Lad os komme i gang med at investere i veje, almene boliger og havvindmøller. Det vil gavne danskerne og den grønne omstilling...

9669 RT @alternativet_: Godt @NWammen og @regeringDK sætter arbejdet i gang med grønne regnemodeller. Det er på høje tid. Nu er det vigtigt, at...

16434 Tillykke med de 18 år @GretaThunberg se\r\n\r\nKampen for klimaet har aldrig været vigtigere. Du er et forbillede for os alle! #dkpoll #dkklima

1142 Godt gået! Glæder mig virkelig over, at vi nu på Fyn er klar til at producere økologisk medicinsk cannabis. Det kan... <https://t.co/pqUemt7T3K>

17657 RT @SamiraNawa: Klimarådet og vismændene er enige. Vi skal have en grøn skattereform nu - på den måde kommer vi igang med den grønne omstilling...

9494 Nej tak til national populistisk selvtilstrækkelighed. De globale udfordringer, klimakrisen og nu corona, viser tydeligt... <https://t.co/o/n5e3lXhA76>

1561 Vi har netop løftet sløret for vores skatteplan, som lægger etter skatter på arbejde, bolig og grønne biler og så vil vi... <https://t.co/o/10yXfRoZxx>

14938 RT @larskohler: Wow!!\r\n\r\nHar ærlig talt lidt frygtet @Spolitik's nye klima- & energiudspil. Har nok med årene mistet lidt af min barmetro på...

7897 RT @alternativet_: STOP KLIMASVIGT!\r\n\r\nRegeringen har længe snakket om at ville gøre noget for klimaet - men det meste er endt i udvalget og sk...

13905 R og SF har med regeringen, V og K udskudt beslutningen om en CO2-afgift i flere år. Ingen garanti for, at der over... <https://t.co/IHcJMQEW5V>

926 RT @ThDanielsen: . @LeaWermelin har helt sikkert hulket af grin, da hun lavede svaret på mit spørgsmål... Helt uden at vide, at klimaafslut...

2975 RT @KlimaMin: Bred opbakning til Danmarks nye klimalov. Stærkt at se samarbejde på tværs af politisk forskellighed, når det gælder vores fr...

12814 RT @Chief_Morse: Der bruges 1,5 kWh for at producere en liter benzin. Så en benzinbil bruger lige så meget strøm som en Tesla. Nå ja, så b...

14891 S vælger en vej med grønne ambitioner, en styrkelse af uddannelse/velfærd/social retfærdighed og en udlændingepolitik, hvo... <https://t.co/o/nb26FgXANK>

6952 @KathrineOllsdag @elbilalliancen Og derfor vi skal have ladeinfrastruktur speedet op ⚡🔌

12504 @carstendyb kjaer @SikandaSIDDIQUE Nemlig. Afgiften skal ligge der hvor klimabelastningen ligger.

12892 RT @jacobmark_sf: Tak til Uffe og Alternativet for at sætte grønne og kulturelle aftryk på finansloven - og for hele tiden at bidrage til d...

10487 RT @DanJoergensen: Jeg har glædet mig til i dag: På trods af en svær situation for DK kommer første del af klimaudspil fra @rege

ringDK: Vi...

11646 @JanEJoergensen Brug fantasien, Jan E: Man kan også lave motion i naturen - feks. vrimler det med mennesker der mo... <https://t.co/uAhYZbTRcr>

17315 ✓Nulafgift på elbiler under 400.000 kr.\r\n✓3 nye havvindmølleparker \r\n✓El bliver 100 % grøn i 2030\r\n✓Lever op til Par... <https://t.co/R52ifZkOu5>

3843 RT @LeaWermelin: Også derfor er regeringens skovfond en god ide. Virksomheder og forbrugere skal have sikkerhed for, at klimakompensation n...

10013 RT @PiaOlsen: Velkommen til SF - Danmarks grønne parti
♥☺☺ Rasmus! Det glæder mig, at vi nu sammen i @SFpolitik skal kæmpe for et grønne o...

1916 RT @UddanFrem: NY SAMRÅDS DATO om #grønneudd - vi glæder os!

6891 Skatten skal være grøn, uden at det bliver dyrere at være dansker. @KonservativeDK er gået i grønt udbrud i blå blo... <https://t.co/Ds8HNRy4y9>

9285 RT @PiaOlsen: ENDNU ET STIK HJEM TIL KLIMAET! Det nytter at bide sig fast i forhandlingsbordet. Vi er tilbage på sporet med et mål om 50-54...

3682 Socialdemokratiet vil oprette en 'grøn fremtidsfond'. Den skal øge de grønne investeringer og samtidig åbne en bagd... <https://t.co/cvuEiQlPVM>

183 Dejligt at Klimarådet også anbefaler 2030 som stop for salg af diesel- og benzinbiler! Vi @KonservativeDK er helt e... <https://t.co/tF44ItDXh>

7137 I går var Vandets Dag. landbrugets sprøjtegifte udvaskes til grundvand og drikkevand i stadig stigende grad.... <https://t.co/XFmFAJjyKQ>

5468 @TroelsBeha Grundpointen i Grønnegaard-bogen er, at omkalfatringen af SKAT er hovedårsagen til miseren. At vi ikke... <https://t.co/aZ00Gcvdnd>

5775 RT @Spolitik: Mette F.: "Danmark skal være en grøn stormagt igen." #dkpol #dkgreen <https://t.co/kbP6s7zA8P>

18047 RT @metteabildgaard: RIGTIG god aftale for det danske miljø! Mere intelligent regulering af landbruget fremover - vil gavne vandmiljøet. Og...

100 @h_kra gesteen @MortenDDHansen 20 mio rækker dælme ikke langt, når det gælder urørt skov. :)

10404 Velskrevet analyse (og klar opbakning) om @Spolitik-forlaget om en grøn fremtidsfond i dagens @politiken. Fonden s... <https://t.co/gIDxmznVR>

111 Dejligt endelig at være i gang med reelle forhandlinger om klimaloven! Jeg forhandler på vegne af @KonservativeDK... <https://t.co/kipVTCHMC8>

14 RT @PiaOlsen: Kære konservative ☹️ Jeg håber I får et godt landsråd! Og at I sætter fut på de grønne ambitioner i blå blok 🙏 @SorenPape @_Ma...

6758 @RasmusJarlov Velfærdslov (som det nok mest vanvittige) \r\n\r\nKlimalov fortsat (regning ukendt - formentlig +30 mia.)... <https://t.co/zanFWANSFL>

7737 DF vil sikre et ordentligt indeklima og forbedre læringsmiljøet i folkeskolen. Håber vi kan få opbakning til vores... <https://t.co/jhi36uPzkU>

4891 @signe_munk Troede det var klimarådets 'model' du ville have? Lidt friskt hvis ikke du har overslag på hvor mange m... <https://t.co/azWqGZHMSl>

7684 Den grønneste regering i nyere tid, vil selvfølgelig ik

ke stemme for en klimalov - i hvert fald ikke med bindende m... <https://t.co/yuYU5W05oF>

17001 Værsgo, Dan Jørgensen: Her har du de mange milliarder, der skal til for at finansiere den grønne omstilling. Lad os... <https://t.co/o/6SRjGHu57k>

11401 RT @emilnielsen: Godt grønt finanslovsudspil fra @radikale, som med 200 millioner kroner årligt i 2021 og 2022 vil sikre mere urørt skov og...

3359 DI er klar med klimaplan. Vil blandt andet elektrificere jernbanenettet og bygge flere havvindmøller. Tak for det... <https://t.co/o/yqDHCbQdVi>

12754 Det er nu, vi skal investere klogt og grønt. Vi skal udnytte ekspertviden, elektrificere Danmark, flere nye vindmøller... <https://t.co/o/geevIVC4QI>

17467 RT @stinuslindgreen: Årets tale fra @oestergaard på #rvm20 kommer omkring coronaepidemien, klimaforandringer, sexisme og flygtningene i M...

4102 @AndreasBachM @spisekammeret @PiaOlsen @PSkipperEL Ifølge Effektiv Landbrug skyder den ansvarlige for demonstration... <https://t.co/o/SUM04pCN16>

6491 RT @oestergaard: Forudsætningen for klimahandling er klimaforhandling. Efter den vådeste februar nogensinde kommer invitationen derfor ikke...

3250 Til fejring af dagen, hvor Folketinget 1. Behandlet L 14 Anlægslov for dobbeltsporede og elektrificerede jernbane o... <https://t.co/o/EsvCEja7SK>

13271 RT @KaareDybvad: Regeringen indgår aftale med DF, EL, SF og Å om indgreb mod udenlandske kapitalfonde: 5-årig karenperiode, skærpet grønt...

1526 Nedslidning er i sagens natur meget alvorligt, men det vil være meget dumt, hvis det ender med at unge dropper at t... <https://t.co/o/aQ858AIHus>

7302 .@enhedslisten er klar til at stemme for forslaget. Når der bruges roundup i landbruget skal den der sprøjter have... <https://t.co/o/jvOnUZodaY>

6436 RT @oestergaard: Siger selvfølgelig en del om os. Og måske lidt om statistikken. Næppe det samme vi og V siger om klima i hvert fald 😊 <https://t.co/o/6SRjGHu57k>

1290 @PeterMa20502201 Vi vil gerne overholde skattestoppet og investere i det grønne. Regeringen er kun pyntegrøn

8149 Helt enig med @aahlers - Øget forskning og innovation er en vigtig del af løsningen på de globale klimaudfordringer... <https://t.co/o/rTMCjT1862>

5378 RT @cyklistforbund: Cykelstier langs statsvejene 🚲 Så vigtigt for at styrke Cykellandet DK. Det er godt for trængslen, sundheden, klimaet....

2466 Klima på finansloven. Skov, flyafgift, og elbiler 🌍 Vi skal have klimahandling nu 🇩🇰 #dkgreen #dkpol <https://t.co/o/MN2eVyl06>

16039 RT @signe_munk: Helt uenig med de økonomiske vismænd. Danmark skal leverer sit fair bidrag til Paris-aftalen. Det gør en klimalo v med 70 %...

7808 Så lige, at Vatikanet vil være CO2 neutral i 2050...Stærkt, Frans... Nu kører det!!!

Name: text, dtype: object

In [7]:

```
# create excel writer object
writer = pd.ExcelWriter('green_sample_full_text.xlsx')
# write dataframe to excel
sample_green.to_excel(writer)
# save the excel
writer.save()
print('DataFrame is written successfully to Excel File.')
```

DataFrame is written successfully to Excel File.

In []:

```
#taking a sample of the most probable words for topic vehicles
sample_vehicles = df.nlargest(100, 'Vehicles')
```

In []:

```
# create excel writer object
writer = pd.ExcelWriter('sample_vehicles.xlsx')
# write dataframe to excel
sample_vehicles.to_excel(writer)
# save the excel
writer.save()
print('DataFrame is written successfully to Excel File.')
```

In []:

```
#taking a sample of the most probable words for agriculture vehicles
sample_agri = df.nlargest(100, 'Agriculture')
```

In []:

```
# create excel writer object
writer = pd.ExcelWriter('sample_agri.xlsx')
# write dataframe to excel
sample_agri.to_excel(writer)
# save the excel
writer.save()
print('DataFrame is written successfully to Excel File.')
```

In []:

```
#taking a sample of the most probable words for topic vehicles
sample_eu = df.nlargest(100, 'European community')
```

In []:

```
# create excel writer object
writer = pd.ExcelWriter('sample_eu.xlsx')
# write dataframe to excel
sample_eu.to_excel(writer)
# save the excel
writer.save()
print('DataFrame is written successfully to Excel File.')
```

1. Installing relevant packages

In []:

```
# install dependencies
!pip install -q condacolab
import condacolab
condacolab.install()

import condacolab
#condacolab.check()

! conda config --add channels conda-forge
! conda config --add channels ostrokach-forge
! conda config --add channels pkgw-forge

! conda install gtk3
! conda install pygobject graph-tool cairo
! conda install -c conda-forge graph-tool
! git clone https://github.com/martingerlach/hSBM_Topicmodel.git
```

```

📄 Downloading https://github.com/jaimergp/miniforge/releases/latest/download/Mambaforge-colab-Linux-x86_64.sh...
📦 Installing...
🔧 Adjusting configuration...
🩹 Patching environment...
🕒 Done in 0:00:35
🔄 Restarting kernel...

```

```
Collecting package metadata (current_repodata.json): - 00\ 00\ 00/ 00- 00\
00\ 00/ 00- 00\ 00\ 00/ 00- 00\ 00\ 00/ 00- 00\ 00\ 00/ 00- 00\ 00\
00- 00\ 00\ 00/ 00- 00\ 00\ 00/ 00- 00\ 00\ 00/ 00- 00\ 00\ 00/ 00- 00\
00\ 00/ 00- 00\ 00\ 00/ 00- 00\ 00\ 00/ 00- 00\ 00\ 00/ 00- 00\
00\ 00/ 00- 00\ 00\ 00/ 00- 00\ 00\ 00/ 00- 00\ 00\ 00/ 00- 00\
00\ 00/ 00- 00\ 00\ 00/ 00- 00\ 00\ 00done
```

```
Solving environment: | 00/ 00- 00\ 00| 00/ 00- 00\ 00| 00/ 00- 00\ 00| 00/
00- 00\ 00| 00/ 00- 00\ 00| 00/ 00- 00\ 00| 00/ 00- 00\ 00| 00/ 00- 00\
00done
```

Package Plan

```
environment location: /usr/local
```

added / updated specs:

- gtk3

The following packages will be downloaded:

	package	build		
range	at-spi2-atk-2.38.0	h0630a04_3	332 KB	conda-forge
range	at-spi2-core-2.40.2	h0630a04_0	647 KB	conda-forge
range	atk-1.0-2.36.0	h3371d22_4	560 KB	conda-forge
range	ca-certificates-2021.5.30	ha878542_0	136 KB	conda-forge
range	cairo-1.16.0	h6cf1ce9_1008	1.5 MB	conda-forge
range	certifi-2021.5.30	py37h89c1867_0	141 KB	conda-forge
range	conda-4.10.1	py37h89c1867_0	3.1 MB	conda-forge
range	dbus-1.13.6	h48d8840_2	572 KB	conda-forge
range	epoxy-1.5.8	h7f98852_0	1.5 MB	conda-forge
range	expat-2.4.1	h9c3ff4c_0	182 KB	conda-forge
-forge	font-ttf-dejavu-sans-mono-2.37	hab24e00_0	388 KB	conda-forge
range	font-ttf-inconsolata-3.000	h77eed37_0	94 KB	conda-forge
-forge	font-ttf-source-code-pro-2.038	h77eed37_0	684 KB	conda-forge
range	font-ttf-ubuntu-0.83	hab24e00_0	1.9 MB	conda-forge
range	fontconfig-2.13.1	hba837de_1005	357 KB	conda-forge
range	fonts-conda-ecosystem-1	0	4 KB	conda-forge

rges

rge

rge

rges

rge

rges

rge

rge

rge

rge

-forge

rge

-forge

rge

rge

rge

fonts-conda-forge-1		0	4 KB	conda-fo
rge				
freetype-2.10.4		h0708190_1	890 KB	conda-fo
rge				
fribidi-1.0.10		h36c2ea0_0	112 KB	conda-fo
rge				
gdk-pixbuf-2.42.6		h04a7f16_0	609 KB	conda-fo
rge				
gettext-0.19.8.1		h0b5b191_1005	3.6 MB	conda-fo
rge				
glib-2.68.3		h9c3ff4c_0	449 KB	conda-fo
rge				
glib-tools-2.68.3		h9c3ff4c_0	86 KB	conda-fo
rge				
graphite2-1.3.13		h58526e2_1001	102 KB	conda-fo
rge				
gtk3-3.24.29		h8879c87_1	9.7 MB	conda-fo
rge				
harfbuzz-2.8.1		h83ec7ef_0	1.9 MB	conda-fo
rge				
hicolor-icon-theme-0.17		ha770c72_2	14 KB	conda-fo
rge				
jpeg-9d		h36c2ea0_0	264 KB	conda-fo
rge				
krb5-1.19.1		hcc1bbae_0	1.4 MB	conda-fo
rge				
libcups-2.3.3		hf5a7f15_0	4.5 MB	conda-fo
rge				
libcurl-7.77.0		h2574ce0_0	334 KB	conda-fo
rge				
libglib-2.68.3		h3e27bee_0	3.1 MB	conda-fo
rge				
libpng-1.6.37		h21135ba_2	306 KB	conda-fo
rge				
libtiff-4.2.0		h85742a9_0	502 KB	
libuuid-2.32.1		h7f98852_1000	28 KB	conda-fo
rge				
libwebp-base-1.2.0		h7f98852_2	815 KB	conda-fo
rge				
libxcb-1.13		h7f98852_1003	395 KB	conda-fo
rge				
openssl-1.1.1k		h7f98852_0	2.1 MB	conda-fo
rge				
pango-1.48.5		hb8ff022_0	398 KB	conda-fo
rge				
pcre-8.44		he1b5a44_0	261 KB	conda-fo
rge				
pixman-0.40.0		h36c2ea0_0	627 KB	conda-fo
rge				
pthread-stubs-0.4		h36c2ea0_1001	5 KB	conda-fo
rge				
xorg-fixesproto-5.0		h7f98852_1002	9 KB	conda-fo
rge				
xorg-inputproto-2.3.2		h7f98852_1002	19 KB	conda-fo
rge				
xorg-kbproto-1.0.7		h7f98852_1002	27 KB	conda-fo
rge				
xorg-libice-1.0.10		h7f98852_0	58 KB	conda-fo
rge				
xorg-libsm-1.2.3		hd9c2040_1000	26 KB	conda-fo
rge				

xorg-libx11-1.7.2		h7f98852_0	941 KB	conda-fo
rge				
xorg-libxau-1.0.9		h7f98852_0	13 KB	conda-fo
rge				
xorg-libxdmcp-1.1.3		h7f98852_0	19 KB	conda-fo
rge				
xorg-libxext-1.3.4		h7f98852_1	54 KB	conda-fo
rge				
xorg-libxfixes-5.0.3		h7f98852_1004	18 KB	conda-fo
rge				
xorg-libxi-1.7.10		h7f98852_0	46 KB	conda-fo
rge				
xorg-libxrender-0.9.10		h7f98852_1003	32 KB	conda-fo
rge				
xorg-libxtst-1.2.3		h7f98852_1002	31 KB	conda-fo
rge				
xorg-recordproto-1.14.2		h7f98852_1002	8 KB	conda-fo
rge				
xorg-renderproto-0.11.1		h7f98852_1002	9 KB	conda-fo
rge				
xorg-xextproto-7.3.0		h7f98852_1002	28 KB	conda-fo
rge				
xorg-xproto-7.0.31		h7f98852_1007	73 KB	conda-fo
rge				

Total:			45.7 MB	

The following NEW packages will be INSTALLED:

at-spi2-atk	conda-forge/linux-64::at-spi2-atk-2.38.0-h0630a04_3
at-spi2-core	conda-forge/linux-64::at-spi2-core-2.40.2-h0630a04_0
atk-1.0	conda-forge/linux-64::atk-1.0-2.36.0-h3371d22_4
cairo	conda-forge/linux-64::cairo-1.16.0-h6cf1ce9_1008
dbus	conda-forge/linux-64::dbus-1.13.6-h48d8840_2
epoxy	conda-forge/linux-64::epoxy-1.5.8-h7f98852_0
expat	conda-forge/linux-64::expat-2.4.1-h9c3ff4c_0
font-ttf-dejavu-s~	conda-forge/noarch::font-ttf-dejavu-sans-mono-2.37-ha
b24e00_0	
font-ttf-inconsol~	conda-forge/noarch::font-ttf-inconsolata-3.000-h77eed
37_0	
font-ttf-source-c~	conda-forge/noarch::font-ttf-source-code-pro-2.038-h7
7eed37_0	
font-ttf-ubuntu	conda-forge/noarch::font-ttf-ubuntu-0.83-hab24e00_0
fontconfig	conda-forge/linux-64::fontconfig-2.13.1-hba837de_1005
fonts-conda-ecossy~	conda-forge/noarch::fonts-conda-ecosystem-1-0
fonts-conda-forge	conda-forge/noarch::fonts-conda-forge-1-0
freetype	conda-forge/linux-64::freetype-2.10.4-h0708190_1
fribidi	conda-forge/linux-64::fribidi-1.0.10-h36c2ea0_0
gdk-pixbuf	conda-forge/linux-64::gdk-pixbuf-2.42.6-h04a7f16_0
gettext	conda-forge/linux-64::gettext-0.19.8.1-h0b5b191_1005
glib	conda-forge/linux-64::glib-2.68.3-h9c3ff4c_0
glib-tools	conda-forge/linux-64::glib-tools-2.68.3-h9c3ff4c_0
graphite2	conda-forge/linux-64::graphite2-1.3.13-h58526e2_1001
gtk3	conda-forge/linux-64::gtk3-3.24.29-h8879c87_1
harfbuzz	conda-forge/linux-64::harfbuzz-2.8.1-h83ec7ef_0
hicolor-icon-theme	conda-forge/linux-64::hicolor-icon-theme-0.17-ha770c7
2_2	
jpeg	conda-forge/linux-64::jpeg-9d-h36c2ea0_0
libcups	conda-forge/linux-64::libcups-2.3.3-hf5a7f15_0
libglib	conda-forge/linux-64::libglib-2.68.3-h3e27bee_0
libpng	conda-forge/linux-64::libpng-1.6.37-h21135ba_2

```

libtiff                pkgs/main/linux-64::libtiff-4.2.0-h85742a9_0
libuuid                conda-forge/linux-64::libuuid-2.32.1-h7f98852_1000
libwebp-base           conda-forge/linux-64::libwebp-base-1.2.0-h7f98852_2
libxcb                 conda-forge/linux-64::libxcb-1.13-h7f98852_1003
pango                  conda-forge/linux-64::pango-1.48.5-hb8ff022_0
pcre                   conda-forge/linux-64::pcre-8.44-he1b5a44_0
pixman                 conda-forge/linux-64::pixman-0.40.0-h36c2ea0_0
pthread-stubs          conda-forge/linux-64::pthread-stubs-0.4-h36c2ea0_1001
xorg-fixesproto        conda-forge/linux-64::xorg-fixesproto-5.0-h7f98852_10
02
xorg-inputproto        conda-forge/linux-64::xorg-inputproto-2.3.2-h7f98852_
1002
xorg-kbproto           conda-forge/linux-64::xorg-kbproto-1.0.7-h7f98852_100
2
xorg-libice            conda-forge/linux-64::xorg-libice-1.0.10-h7f98852_0
xorg-libsm             conda-forge/linux-64::xorg-libsm-1.2.3-hd9c2040_1000
xorg-libx11            conda-forge/linux-64::xorg-libx11-1.7.2-h7f98852_0
xorg-libxau            conda-forge/linux-64::xorg-libxau-1.0.9-h7f98852_0
xorg-libxdmcp          conda-forge/linux-64::xorg-libxdmcp-1.1.3-h7f98852_0
xorg-libxext           conda-forge/linux-64::xorg-libxext-1.3.4-h7f98852_1
xorg-libxfixes         conda-forge/linux-64::xorg-libxfixes-5.0.3-h7f98852_1
004
xorg-libxi             conda-forge/linux-64::xorg-libxi-1.7.10-h7f98852_0
xorg-libxrender         conda-forge/linux-64::xorg-libxrender-0.9.10-h7f98852
_1003
xorg-libxtst           conda-forge/linux-64::xorg-libxtst-1.2.3-h7f98852_100
2
xorg-recordproto       conda-forge/linux-64::xorg-recordproto-1.14.2-h7f9885
2_1002
xorg-renderproto       conda-forge/linux-64::xorg-renderproto-0.11.1-h7f9885
2_1002
xorg-xextproto         conda-forge/linux-64::xorg-xextproto-7.3.0-h7f98852_1
002
xorg-xproto            conda-forge/linux-64::xorg-xproto-7.0.31-h7f98852_100
7

```

The following packages will be UPDATED:

```

ca-certificates        2020.12.5-ha878542_0 --> 2021.5.30-
ha878542_0
certifi                2020.12.5-py37h89c1867_1 --> 2021.5.30-
py37h89c1867_0
conda                  4.9.2-py37h89c1867_0 --> 4.10.1-py3
7h89c1867_0
krb5                   1.17.2-h926e7f8_0 --> 1.19.1-hcc
1bbae_0
libcurl                7.75.0-hc4aaa36_0 --> 7.77.0-h25
74ce0_0
openssl                1.1.1j-h7f98852_0 --> 1.1.1k-h7f
98852_0

```

Downloading and Extracting Packages

```

harfbuzz-2.8.1        | 1.9 MB      | : 100% 1.0/1 [00:00<00:00, 1.66it/s]
libwebp-base-1.2.0    | 815 KB      | : 100% 1.0/1 [00:00<00:00, 4.39it/s]
pango-1.48.5          | 398 KB      | : 100% 1.0/1 [00:00<00:00, 6.80it/s]
libpng-1.6.37         | 306 KB      | : 100% 1.0/1 [00:00<00:00, 8.64it/s]
ca-certificates-2021  | 136 KB      | : 100% 1.0/1 [00:00<00:00, 13.95it/s]
font-ttf-dejavu-sans  | 388 KB      | : 100% 1.0/1 [00:00<00:00, 10.11it/s]
font-ttf-source-code  | 684 KB      | : 100% 1.0/1 [00:00<00:00, 5.14it/s]

```

libtiff-4.2.0	502 KB	: 100% 1.0/1 [00:00<00:00, 5.13it/s]
font-ttf-ubuntu-0.83	1.9 MB	: 100% 1.0/1 [00:00<00:00, 2.73it/s]
fribidi-1.0.10	112 KB	: 100% 1.0/1 [00:00<00:00, 15.60it/s]
xorg-libxau-1.0.9	13 KB	: 100% 1.0/1 [00:00<00:00, 21.07it/s]
xorg-libxfixed-5.0.3	18 KB	: 100% 1.0/1 [00:00<00:00, 25.64it/s]
xorg-kbproto-1.0.7	27 KB	: 100% 1.0/1 [00:00<00:00, 25.25it/s]
pthread-stubs-0.4	5 KB	: 100% 1.0/1 [00:00<00:00, 22.94it/s]
libxcb-1.13	395 KB	: 100% 1.0/1 [00:00<00:00, 5.24it/s]
xorg-libxext-1.3.4	54 KB	: 100% 1.0/1 [00:00<00:00, 17.85it/s]
conda-4.10.1	3.1 MB	: 100% 1.0/1 [00:00<00:00, 1.21it/s]
xorg-libice-1.0.10	58 KB	: 100% 1.0/1 [00:00<00:00, 19.52it/s]
xorg-libsm-1.2.3	26 KB	: 100% 1.0/1 [00:00<00:00, 22.07it/s]
libglib-2.68.3	3.1 MB	: 100% 1.0/1 [00:00<00:00, 1.28it/s]
font-ttf-inconsolata	94 KB	: 100% 1.0/1 [00:00<00:00, 19.15it/s]
fonts-conda-forge-1	4 KB	: 100% 1.0/1 [00:00<00:00, 26.02it/s]
epoxy-1.5.8	1.5 MB	: 100% 1.0/1 [00:00<00:00, 2.69it/s]
openssl-1.1.1k	2.1 MB	: 100% 1.0/1 [00:00<00:00, 2.32it/s]
xorg-fixedproto-5.0	9 KB	: 100% 1.0/1 [00:00<00:00, 22.25it/s]
jpeg-9d	264 KB	: 100% 1.0/1 [00:00<00:00, 8.13it/s]
pixman-0.40.0	627 KB	: 100% 1.0/1 [00:00<00:00, 4.26it/s]
fontconfig-2.13.1	357 KB	: 100% 1.0/1 [00:00<00:00, 7.88it/s]
xorg-xproto-7.0.31	73 KB	: 100% 1.0/1 [00:00<00:00, 17.07it/s]
atk-1.0-2.36.0	560 KB	: 100% 1.0/1 [00:00<00:00, 4.51it/s]
expat-2.4.1	182 KB	: 100% 1.0/1 [00:00<00:00, 13.64it/s]
hicolor-icon-theme-0	14 KB	: 100% 1.0/1 [00:00<00:00, 21.28it/s]
xorg-libxdmcp-1.1.3	19 KB	: 100% 1.0/1 [00:00<00:00, 24.12it/s]
freetype-2.10.4	890 KB	: 100% 1.0/1 [00:00<00:00, 4.40it/s]
dbus-1.13.6	572 KB	: 100% 1.0/1 [00:00<00:00, 5.65it/s]
xorg-libxi-1.7.10	46 KB	: 100% 1.0/1 [00:00<00:00, 20.14it/s]
at-spi2-core-2.40.2	647 KB	: 100% 1.0/1 [00:00<00:00, 5.57it/s]
xorg-libxrender-0.9.	32 KB	: 100% 1.0/1 [00:00<00:00, 23.19it/s]
glib-2.68.3	449 KB	: 100% 1.0/1 [00:00<00:00, 4.46it/s]
xorg-libxtst-1.2.3	31 KB	: 100% 1.0/1 [00:00<00:00, 22.19it/s]
libcurl-7.77.0	334 KB	: 100% 1.0/1 [00:00<00:00, 10.19it/s]
certifi-2021.5.30	141 KB	: 100% 1.0/1 [00:00<00:00, 14.80it/s]
gtk3-3.24.29	9.7 MB	: 100% 1.0/1 [00:02<00:00, 2.77s/it]
xorg-xextproto-7.3.0	28 KB	: 100% 1.0/1 [00:00<00:00, 21.24it/s]
fonts-conda-ecosyste	4 KB	: 100% 1.0/1 [00:00<00:00, 26.69it/s]
glib-tools-2.68.3	86 KB	: 100% 1.0/1 [00:00<00:00, 15.52it/s]
graphite2-1.3.13	102 KB	: 100% 1.0/1 [00:00<00:00, 18.18it/s]
gettext-0.19.8.1	3.6 MB	: 100% 1.0/1 [00:01<00:00, 1.14s/it]
cairo-1.16.0	1.5 MB	: 100% 1.0/1 [00:00<00:00, 2.47it/s]
at-spi2-atk-2.38.0	332 KB	: 100% 1.0/1 [00:00<00:00, 10.41it/s]
libcups-2.3.3	4.5 MB	: 100% 1.0/1 [00:00<00:00, 1.37it/s]
pcre-8.44	261 KB	: 100% 1.0/1 [00:00<00:00, 10.07it/s]
xorg-recordproto-1.1	8 KB	: 100% 1.0/1 [00:00<00:00, 27.27it/s]
xorg-libx11-1.7.2	941 KB	: 100% 1.0/1 [00:00<00:00, 4.29it/s]
libuuid-2.32.1	28 KB	: 100% 1.0/1 [00:00<00:00, 23.12it/s]
xorg-inputproto-2.3.	19 KB	: 100% 1.0/1 [00:00<00:00, 25.01it/s]
xorg-renderproto-0.1	9 KB	: 100% 1.0/1 [00:00<00:00, 24.32it/s]
krb5-1.19.1	1.4 MB	: 100% 1.0/1 [00:00<00:00, 2.36it/s]
gdk-pixbuf-2.42.6	609 KB	: 100% 1.0/1 [00:00<00:00, 3.61it/s]

Preparing transaction: / 00- 00\ 00done

Verifying transaction: / 00- 00\ 00| 00/ 00- 00\ 00| 00/ 00- 00\ 00| 00/
00- 00\ 00| 00/ 00- 00\ 00done

Executing transaction: / 00- 00\ 00| 00/ 00- 00\

00| 00/ 00- 00\

00| 00/ 00- 00done

Collecting package metadata (current_repodata.json): - 00\ 00| 00/ 00- 00\
00| 00/ 00- 00\ 00| 00/ 00- 00\ 00| 00/ 00- 00\ 00| 00/ 00- 00\ 00| 00/
00- 00\ 00| 00/ 00- 00done

Solving environment: | 00/ 00- 00\ 00| 00/ 00- 00\ 00| 00/ 00- 00\ 00| 00/
00- 00\ 00| 00/ 00- 00\ 00| 00/ 00- 00\ 00| 00/ 00- 00\ 00| 00/ 00- 00\
00| 00/ 00- 00\ 00| 00/ 00- 00\ 00| 00/ 00- 00\ 00| 00/ 00- 00\ 00| 00/
00- 00\ 00| 00/ 00- 00\ 00| 00/ 00- 00\ 00| 00/ 00- 00\ 00| 00/ 00- 00\
00| 00/ 00- 00\ 00| 00/ 00- 00\ 00| 00/ 00- 00\ 00| 00/ 00- 00\ 00| 00/
00- 00\ 00| 00/ 00- 00\ 00| 00/ 00- 00\ 00| 00/ 00- 00\ 00| 00/ 00done

Package Plan

environment location: /usr/local

added / updated specs:

- cairo
- graph-tool
- pygobject

The following packages will be downloaded:

package	build			
-----	-----			
boost-1.74.0	py37h6dcda5c_3	342 KB	conda-fo	
rge				
boost-cpp-1.74.0	hc6e9bd1_3	16.3 MB	conda-fo	
rge				
cairomm-1.12.2	ha770c72_3	11 KB	conda-fo	
rge				
cairomm-1.0-1.12.2	h56b4340_3	690 KB	conda-fo	
rge				
cyclcr-0.10.0	py_2	9 KB	conda-fo	
rge				
gmp-6.2.1	h58526e2_0	806 KB	conda-fo	
rge				
gobject-introspection-1.68.0	py37hb87790b_1	1.5 MB	conda-f	
orge				
graph-tool-2.37	py37h6056006_1	76.4 MB	conda-fo	
rge				
kiwisolver-1.3.1	py37h2527ec5_1	78 KB	conda-fo	
rge				
lcms2-2.12	hddcbb42_0	443 KB	conda-fo	
rge				
libblas-3.9.0	9_openblas	11 KB	conda-fo	
rge				
libcbblas-3.9.0	9_openblas	11 KB	conda-fo	
rge				
libgfortran-ng-9.3.0	hff62375_19	22 KB	conda-fo	
rge				
libgfortran5-9.3.0	hff62375_19	2.0 MB	conda-fo	
rge				
liblapack-3.9.0	9_openblas	11 KB	conda-fo	
rge				
libopenblas-0.3.15	pthread_h8fe5266_1	9.2 MB	conda-	
forge				
librsvg-2.50.7	hc3c00ef_0	17.0 MB	conda-fo	
rge				
libxml2-2.9.12	h72842e0_0	772 KB	conda-fo	
rge				
matplotlib-base-3.4.2	py37hdd32ed1_0	7.2 MB	conda-fo	
rge				
numpy-1.20.3	py37h038b26d_1	5.7 MB	conda-fo	
rge				

olefile-0.46		pyh9f0ad1d_1	32 KB	conda-fo
rge				
openjpeg-2.4.0		hb52868f_1	444 KB	conda-fo
rge				
pillow-8.2.0		py37h4600e1f_1	684 KB	conda-fo
rge				
pycairo-1.20.1		py37hfff247e_0	77 KB	conda-fo
rge				
pygobject-3.40.1		py37he6fa580_1	286 KB	conda-fo
rge				
pyparsing-2.4.7		pyh9f0ad1d_0	60 KB	conda-fo
rge				
python-dateutil-2.8.1		py_0	220 KB	conda-fo
rge				
scipy-1.6.3		py37h29e03ee_0	20.5 MB	conda-fo
rge				
sigcpp-2.0-2.10.3		h58526e2_0	4.3 MB	conda-fo
rge				
sparsehash-2.0.3		h6bb024c_0	85 KB	ostrokac
h-forge				
tornado-6.1		py37h5e8e339_1	646 KB	conda-fo
rge				
zstandard-0.15.2		py37h5e8e339_0	617 KB	conda-fo
rge				


Total:			166.2 MB	

The following NEW packages will be INSTALLED:

boost	conda-forge/linux-64::boost-1.74.0-py37h6dcda5c_3
boost-cpp	conda-forge/linux-64::boost-cpp-1.74.0-hc6e9bd1_3
cairomm	conda-forge/linux-64::cairomm-1.12.2-ha770c72_3
cairomm-1.0	conda-forge/linux-64::cairomm-1.0-1.12.2-h56b4340_3
cycler	conda-forge/noarch::cycler-0.10.0-py_2
gmp	conda-forge/linux-64::gmp-6.2.1-h58526e2_0
gobject-introspect~	conda-forge/linux-64::gobject-introspection-1.68.0-py
37hb87790b_1	
graph-tool	conda-forge/linux-64::graph-tool-2.37-py37h6056006_1
kiwisolver	conda-forge/linux-64::kiwisolver-1.3.1-py37h2527ec5_1
lcms2	conda-forge/linux-64::lcms2-2.12-hddcbb42_0
libblas	conda-forge/linux-64::libblas-3.9.0-9_openblas
libcbblas	conda-forge/linux-64::libcbblas-3.9.0-9_openblas
libgfortran-ng	conda-forge/linux-64::libgfortran-ng-9.3.0-hff62375_1
9	
libgfortran5	conda-forge/linux-64::libgfortran5-9.3.0-hff62375_19
liblapack	conda-forge/linux-64::liblapack-3.9.0-9_openblas
libopenblas	conda-forge/linux-64::libopenblas-0.3.15-pthreads_h8f
e5266_1	
librsvg	conda-forge/linux-64::librsvg-2.50.7-hc3c00ef_0
matplotlib-base	conda-forge/linux-64::matplotlib-base-3.4.2-py37hdd32
ed1_0	
numpy	conda-forge/linux-64::numpy-1.20.3-py37h038b26d_1
olefile	conda-forge/noarch::olefile-0.46-pyh9f0ad1d_1
openjpeg	conda-forge/linux-64::openjpeg-2.4.0-hb52868f_1
pillow	conda-forge/linux-64::pillow-8.2.0-py37h4600e1f_1
pycairo	conda-forge/linux-64::pycairo-1.20.1-py37hfff247e_0
pygobject	conda-forge/linux-64::pygobject-3.40.1-py37he6fa580_1
pyparsing	conda-forge/noarch::pyparsing-2.4.7-pyh9f0ad1d_0
python-dateutil	conda-forge/noarch::python-dateutil-2.8.1-py_0
scipy	conda-forge/linux-64::scipy-1.6.3-py37h29e03ee_0
sigcpp-2.0	conda-forge/linux-64::sigcpp-2.0-2.10.3-h58526e2_0

Cloning into 'hSBM Topicmodel'...

```
remote: Enumerating objects: 336, done.
remote: Counting objects: 100% (60/60), done.
remote: Compressing objects: 100% (39/39), done.
remote: Total 336 (delta 33), reused 27 (delta 20), pack-reused 276
Receiving objects: 100% (336/336), 22.02 MiB | 23.34 MiB/s, done.
Resolving deltas: 100% (93/93), done.
```



In []:

```
import graph_tool.all as gt
```

In []:

```
from hSBM_Topicmodel.sbmtm import sbmtm
```

2. Importing the data

In []:

```
# access files in google drive
from google.colab import drive
drive.mount('/content/gdrive')
```

Mounted at /content/gdrive

In []:

```
# Load dataset
import pandas as pd
path2tox_data = '/content/gdrive/My Drive/FINAL_DF_GREEN_WITH_PARTI_AFF.csv' # change t
his path to where you put the file
df = pd.read_csv(path2tox_data)
```

In []:

```
df.shape
```

Out[]:

```
(26052, 12)
```

3. Cleaning and preparing the data

In []:

```
#droppping nan values
df = df.dropna(axis=0, subset=('green_words', ))
```

In []:

```
pd.set_option('display.max_colwidth',1000)
#pd.set_option('display.max_rows',1000)
```

In []:

```
df.head(5)
```

Out[]:

	Unnamed: 0	Unnamed: 0.1	Unnamed: 0.1.1	Unnamed: 0.1.1.1	id	created_at	
0	0	5068	7155	7155.0	1387387662387105793	2021-04-28 12:46:09	I Danmark parametre land sammen o
1	2	5083	7170	7170.0	1380080544990519297	2021-04-08 08:50:17	https:// Stærk fæ @NNI omstillir gøre de http
2	3	5090	7177	7177.0	1333500247021146116	2020-11-30 19:56:48	@Dansk erhvervsklir tillykke til https:
3	4	5122	7209	7209.0	1098702372417798146	2019-02-21 21:53:53	Blandt ande dansk l nec https:
4	5	5133	7220	7220.0	1055881691225604097	2018-10-26 17:59:47	@jesp Svæ umiddelbar udsendelse

In []:

```
#making sure that all text is in strings
df['lemma_tweet']=df['lemma_tweet'].apply(str)
```

In []:

```
#replacing nan values with empty string
df.lemma_tweet = df.lemma_tweet.apply(lambda x: '' if str(x) == 'nan' else x)
```

In []:

```
df.shape
```

Out[]:

```
(26052, 12)
```

4. Running the HSBM model

In []:

```
import nltk
tokenizer = nltk.tokenize.casual.TweetTokenizer()
```

In []:

```
#preparing the data for the HSBM model
df['doc'] = df.lemma_tweet.apply(tokenizer.tokenize).values
docs = df['doc']
```

In []:

```
# remove infrequent words.
from collections import Counter
cutoff = 5
c = Counter()
for doc in docs:
    c.update(Counter(doc))
vocab = c.most_common(25000)
vocab = set([w for w, count in vocab if count > 5])
# remove words
docs = [[w for w in doc if w in vocab] for doc in docs]
```

In []:

```
#defining the model
model = sbmtm()
model.make_graph(docs, documents=['%d'%i for i in range(len(docs))])
```

In []:

```
#We define a seed in order to obtain the same model everytime we run it
gt.seed_rng(32)
model.fit()
```

In []:

```
#printing out the topics of the hsbm  
#l = the level which we want the topics on. Level 0 is the most fundamental, level 1 more aggregated  
#n = the n-largest topic loading words we want to see  
model.topics(l=0,n=10)
```

```
/content/hSBM_Topicmodel/sbmtm.py:634: RuntimeWarning: invalid value encountered in true_divide
    p_td_d = (n_db/np.sum(n_db,axis=1)[: ,np.newaxis]).T
/content/hSBM_Topicmodel/sbmtm.py:640: RuntimeWarning: invalid value encountered in true_divide
    p_tw_d = (n_dbw/np.sum(n_dbw,axis=1)[: ,np.newaxis]).T
```

Out[]:

```
{0: [('danmark', 0.9349667271627344),
      ('forrest', 0.034482758620689655),
      ('førertrøje', 0.018753781004234724),
      ('her', 0.011796733212341199)],
 1: [('mange', 0.27047712099006677),
      ('meget', 0.16560820713238886),
      ('lille', 0.12489822504478099),
      ('langt', 0.08891060087933561),
      ('kommune', 0.05829669434945449),
      ('markant', 0.05031753786028334),
      ('gammel', 0.035499104380394074),
      ('ros', 0.030288226673180263),
      ('spare', 0.019866471258752644),
      ('fremme', 0.018726591760299626)],
 2: [('job', 0.1865740740740741),
      ('hjælp', 0.06851851851851852),
      ('miste', 0.0375),
      ('politi', 0.03333333333333333),
      ('tusindvis', 0.031018518518518518),
      ('trussel', 0.02361111111111111),
      ('tryghed', 0.020833333333333332),
      ('arbejdsmarked', 0.020370370370370372),
      ('sundhedsvæsen', 0.017592592592592594),
      ('velfærdssamfund', 0.017592592592592594)],
 3: [('verden', 0.1800037306472673),
      ('land', 0.1680656593919045),
      ('vise', 0.14530871106136914),
      ('hel', 0.136168625256482),
      ('løsning', 0.12497668345457937),
      ('finde', 0.08282036933407946),
      ('rest', 0.037119940309643724),
      ('fattig', 0.023689610147360568),
      ('lokal', 0.019585898153329603),
      ('erfaring', 0.013243797798918112)],
 4: [('landbrug', 0.45383324006715164),
      ('økologi', 0.07330721880246223),
      ('økologisk', 0.06603245663122552),
      ('fødevare', 0.06323447118074986),
      ('landmand', 0.04280917739227756),
      ('plantebasere', 0.03273642977056519),
      ('landbrugspakke', 0.01846670397313934),
      ('kø', 0.01650811415780638),
      ('mark', 0.013989927252378288),
      ('fordoble', 0.012870733072188025)],
 5: [('lade', 0.2656756756756757),
      ('stoppe', 0.12837837837837837),
      ('stop', 0.07324324324324324),
      ('jord', 0.061621621621621624),
      ('droppe', 0.039729729729729726),
      ('slut', 0.02837837837837838),
      ('fremtidig', 0.027837837837837838),
      ('hente', 0.025135135135135135),
      ('forurening', 0.024324324324324326),
      ('tydeligt', 0.02216216216216216)],
 6: [('dansk', 0.4150695720661591),
      ('bæredygtig', 0.17563665003938042),
      ('udvikling', 0.08348647939091625),
      ('udvikle', 0.04568128117616172),
      ('besøge', 0.029666579154633763),
```

('eksport', 0.02520346547650302),
 ('bidrag', 0.022578104489367288),
 ('understøtte', 0.018114990811236544),
 ('afvikle', 0.017327382515095826),
 ('partnerskab', 0.01680231031766868)],
 7: [('stor', 0.2686297438078027),
 ('mere', 0.25757414691187414),
 ('brug', 0.08323588816838524),
 ('foreslå', 0.05549059211225683),
 ('udspil', 0.050706920378441586),
 ('skridt', 0.050706920378441586),
 ('hjælpe', 0.04432869140002126),
 ('mindst', 0.04082066546189008),
 ('masse', 0.027001169341979377),
 ('opbakning', 0.023386839587541192)],
 8: [('tale', 0.07791929787006315),
 ('miljø', 0.07663491383923793),
 ('politisk', 0.07246066573905598),
 ('tid', 0.0713903457133683),
 ('handle', 0.0683934496414428),
 ('læse', 0.053408969281815266),
 ('kamp', 0.045702665096863965),
 ('høre', 0.033501016804024404),
 ('flot', 0.027507224660173392),
 ('ting', 0.022262656534303756)],
 9: [('stærk', 0.21249252839210997),
 ('stemme', 0.16826060968320383),
 ('retning', 0.11506276150627615),
 ('europa', 0.07232516437537358),
 ('fremad', 0.06455469216975493),
 ('grønnere', 0.054094441123729825),
 ('fællesskab', 0.05140466228332337),
 ('friere', 0.02958756724447101),
 ('kandidat', 0.019127316198445904),
 ('grønner', 0.01852958756724447)],
 10: [('klima', 0.6025286448044251),
 ('fælles', 0.06637692611615963),
 ('klimakamp', 0.053733702094033976),
 ('grad', 0.04938759383642829),
 ('håb', 0.035954168312919796),
 ('resultat', 0.03536151718688266),
 ('tværs', 0.019557487159225603),
 ('lederskab', 0.018569735282497037),
 ('stm', 0.017186882655077045),
 ('savne', 0.017186882655077045)],
 11: [('god', 0.2678732352057675),
 ('tak', 0.11167017122258936),
 ('sammen', 0.08320817062180835),
 ('kæmpe', 0.08275758486031841),
 ('arbejde', 0.07464704115349954),
 ('fremtid', 0.07337038149594473),
 ('glad', 0.036046860919194956),
 ('indsats', 0.028462000600781016),
 ('fantastisk', 0.02613397416641634),
 ('stolt', 0.024181435866626615)],
 12: [('grøn', 0.7497683700377735),
 ('omstilling', 0.1691967785617561),
 ('investering', 0.03898510441165989),
 ('investere', 0.024374599101988453),
 ('snak', 0.011973487278169766),
 ('grønbetydneru', 0.00570166060865227)],

13: [('gøre', 0.2939440443867312),
('gå', 0.17766497461928935),
('del', 0.072010388383898),
('selvfølgelig', 0.048518474796364064),
('gerne', 0.04710187699209066),
('nødvendig', 0.043914531932475506),
('tilbage', 0.03683154291110849),
('løfte', 0.031165151694014875),
('trække', 0.02490851139180734),
('al', 0.024672411757761775)],

14: [('dag', 0.32311834850621873),
('være', 0.11770739838440826),
('støtte', 0.0715476343120913),
('dkmedie', 0.06603410693678677),
('skrive', 0.06385434030003846),
('presse', 0.03923579946146942),
('klimaminister', 0.030901397615078856),
('nemlig', 0.024233876137966407),
('grund', 0.02320810360302603),
('melde', 0.016155917425310938)],

15: [('aften', 0.11175059952038369),
('københavn', 0.10167865707434053),
('dagsord', 0.05275779376498801),
('behandle', 0.04460431654676259),
('ordfører', 0.044124700239808155),
('totalt', 0.03405275779376499),
('kritisk', 0.03261390887290168),
('kbh', 0.027817745803357313),
('borgmester', 0.026378896882494004),
('aarhus', 0.025899280575539568)],

16: [('glæde', 0.26557799742157284),
('samarbejde', 0.22346368715083798),
('tillykke', 0.0923936398796734),
('velkommen', 0.0446927374301676),
('sej', 0.04383326171035668),
('dialog', 0.02965191233347658),
('fornøjelse', 0.028792436613665665),
('ære', 0.021057155135367427),
('post', 0.015040825096691019),
('glædelig', 0.015040825096691019)],

17: [('☺', 0.1231209735146743),
('D', 0.09090909090909091),
('K', 0.08732999284180387),
('☹', 0.06442376521116679),
('nordisk', 0.052254831782390834),
('☺', 0.049391553328561204),
('☹', 0.043664996420901936),
('☹', 0.03793843951324266),
('E', 0.02863278453829635),
('U', 0.024337866857551897)],

18: [('håbe', 0.07566789313709807),
('klimahandling', 0.06942889137737962),
('morgen', 0.05855063189889618),
('politiker', 0.0532714765637498),
('møde', 0.0518317069268917),
('spændende', 0.038713805791073426),
('huske', 0.03775395936650136),
('uge', 0.03327467605183171),
('rykke', 0.03199488081906895),
('bakke', 0.029595264757638778)],

19: [('dkenergi', 0.13568464730290455),

('varm', 0.06265560165975104),
 ('klimaforhandling', 0.056016597510373446),
 ('klimaudspil', 0.04937759336099585),
 ('rolle', 0.04771784232365145),
 ('spille', 0.04439834024896266),
 ('svær', 0.043568464730290454),
 ('varmepumper', 0.03900414937759336),
 ('fjernvarme', 0.03900414937759336),
 ('tilskud', 0.025726141078838173)],
 20: [('vigtig', 0.3755868544600939),
 ('fokus', 0.19212712170458648),
 ('kontrol', 0.05055976886962803),
 ('kvinde', 0.0288912964969303),
 ('råbe', 0.026724449259660527),
 ('ligestilling', 0.02239075478512098),
 ('udbud', 0.02239075478512098),
 ('emne', 0.02239075478512098),
 ('målrettet', 0.022029613578909354),
 ('forbedring', 0.020223907547851208)],
 21: [('tænke', 0.14587242026266417),
 ('regne', 0.04643527204502814),
 ('begrænse', 0.039399624765478425),
 ('midt', 0.03564727954971857),
 ('undskyldning', 0.0323639774859287),
 ('forsøg', 0.02626641651031895),
 ('absurd', 0.024859287054409006),
 ('overholde', 0.024859287054409006),
 ('hidtil', 0.024390243902439025),
 ('beregning', 0.021575984990619138)],
 22: [('sige', 0.12278374982549212),
 ('kun', 0.05821583135557727),
 ('mene', 0.041044255200335056),
 ('tro', 0.029945553539019964),
 ('sidste', 0.026036576853273766),
 ('område', 0.022127600167527572),
 ('altså', 0.020661733910372748),
 ('ret', 0.01947508027362837),
 ('synes', 0.018358229791986598),
 ('forstå', 0.018148820326678767)],
 23: [('kl', 0.05904196063869291),
 ('kollega', 0.04455997029335314),
 ('idé', 0.032305978462681025),
 ('dygtig', 0.030449313033791312),
 ('søge', 0.026735982176011884),
 ('mødes', 0.025993316004455997),
 ('dagsorden', 0.02376531748978834),
 ('skarp', 0.021537318975120682),
 ('live', 0.01968065354623097),
 ('fredag', 0.019309320460453028)],
 24: [('energi', 0.3205013428827216),
 ('foregangsland', 0.07699194270367055),
 ('vedvarende', 0.07341092211280215),
 ('energieffektivitet', 0.04028648164726947),
 ('energiudspil', 0.03849597135183527),
 ('vind', 0.03715308863025962),
 ('fastholde', 0.03581020590868397),
 ('dække', 0.02954341987466428),
 ('sol', 0.026410026857654433),
 ('elafgift', 0.021933751119068933)],
 25: [('budskab', 0.07730182176267848),
 ('bæredygtighed', 0.0758247168882324),

('christiansborg', 0.06745445593303791),
 ('konference', 0.05120630231413097),
 ('dele', 0.03692762186115214),
 ('skøn', 0.029542097488921712),
 ('torsdag', 0.028557360905957656),
 ('modtage', 0.026095519448547513),
 ('klimastrejke', 0.024618414574101428),
 ('engagere', 0.023633677991137372)],
 26: [('skade', 0.06886858749121574),
 ('mink', 0.05832747716092762),
 ('biogas', 0.05340829234012649),
 ('etablere', 0.04356992269852424),
 ('forudsætning', 0.035839775122979624),
 ('vandløb', 0.02881236823612087),
 ('død', 0.025298664792691498),
 ('tjene', 0.024595924104005622),
 ('frisk', 0.023893183415319746),
 ('desuden', 0.023190442726633873)],
 27: [('så', 0.1633516401959329),
 ('komme', 0.08627727475137302),
 ('se', 0.05941071693632181),
 ('helt', 0.050319133145316906),
 ('give', 0.04764732076591955),
 ('gang', 0.03940923259611103),
 ('lige', 0.03272970164761763),
 ('bare', 0.032358616594923555),
 ('forslag', 0.029389936173370938),
 ('enig', 0.028759091583791006)],
 28: [('menneske', 0.18560277536860365),
 ('ung', 0.18256721595836947),
 ('liv', 0.09800520381613183),
 ('konsekvens', 0.0745880312228968),
 ('naturlig', 0.05941023417172593),
 ('forbedre', 0.04640069384215091),
 ('rettighed', 0.035993061578490894),
 ('øje', 0.03252385082393756),
 ('optage', 0.02818733738074588),
 ('viden', 0.02515177797051171)],
 29: [('forskning', 0.2718064153066967),
 ('ekstra', 0.10635903207653348),
 ('massivt', 0.04952166572875633),
 ('bedst', 0.03432751828925155),
 ('innovation', 0.032639279684862126),
 ('satsning', 0.032076533483398985),
 ('holland', 0.023635340461451885),
 ('tilfreds', 0.023635340461451885),
 ('dertil', 0.020258863252673044),
 ('pulje', 0.019696117051209903)],
 30: [('natur', 0.3303988261971455),
 ('biodiversitet', 0.07362945178071229),
 ('plads', 0.060157396291850074),
 ('vild', 0.04668534080298786),
 ('beskytte', 0.029611844737895156),
 ('naturnationalpark', 0.026277177537681738),
 ('dknatur', 0.024409763905562223),
 ('areal', 0.023209283713485393),
 ('true', 0.022408963585434174),
 ('plante', 0.01854074963318661)],
 31: [('følge', 0.12589473684210525),
 ('sidde', 0.055578947368421054),
 ('hjem', 0.032842105263157895),

('tyskland', 0.031578947368421054),
 ('mandag', 0.03031578947368421),
 ('samtale', 0.026947368421052633),
 ('skyde', 0.026105263157894736),
 ('link', 0.023578947368421053),
 ('minde', 0.019789473684210527),
 ('styre', 0.01768421052631579)],
 32: [('velfærd', 0.31626666666666664),
 ('venstre', 0.25866666666666666),
 ('social', 0.14186666666666667),
 ('reform', 0.0336),
 ('nedskæring', 0.028266666666666666),
 ('retfærdighed', 0.02666666666666667),
 ('lønmodtager', 0.0256),
 ('dumping', 0.025066666666666668),
 ('ærgeligt', 0.024),
 ('bnp', 0.019733333333333332)],
 33: [('økonomi', 0.16730917501927525),
 ('økonomisk', 0.1661526599845798),
 ('krise', 0.09020817270624518),
 ('ansvarlig', 0.06399383191981496),
 ('genstart', 0.0551272166538165),
 ('genopretning', 0.04934464148033924),
 ('langsigtet', 0.028141865844255976),
 ('solidarisk', 0.027756360832690823),
 ('hjælpepak', 0.020046260601387818),
 ('kickstarte', 0.016962220508866616)],
 34: [('klimapolitik', 0.18963675213675213),
 ('klimapartnerskab', 0.07745726495726496),
 ('lancere', 0.057692307692307696),
 ('pege', 0.04861111111111111),
 ('løbe', 0.03205128205128205),
 ('realistisk', 0.030982905982905984),
 ('spids', 0.028311965811965812),
 ('forsvar', 0.027243589743589744),
 ('dronning', 0.026709401709401708),
 ('landbrugsforhandling', 0.023504273504273504)],
 35: [('klimalov', 0.33577586206896554),
 ('binde', 0.1254310344827586),
 ('vedtage', 0.07887931034482759),
 ('borgerforslag', 0.06853448275862069),
 ('reduktionsmål', 0.04008620689655172),
 ('forpligtede', 0.03405172413793103),
 ('opposition', 0.03146551724137931),
 ('folketingssal', 0.027155172413793102),
 ('jul', 0.01939655172413793),
 ('ommer', 0.014224137931034483)],
 36: [('aftale', 0.6468551644547028),
 ('bred', 0.2423542989036353),
 ('indgå', 0.10444316214656665),
 ('off', 0.006347374495095211)],
 37: [('klimaråd', 0.24173748819641172),
 ('løkke', 0.12370160528800755),
 ('rapport', 0.1052880075542965),
 ('formand', 0.10056657223796034),
 ('kritik', 0.0684608120868744),
 ('lars', 0.061378659112370164),
 ('vurdering', 0.024551463644948063),
 ('fyre', 0.024079320113314446),
 ('peter', 0.0226628895184136),
 ('konklusion', 0.019830028328611898)],

38: [('ambitiøst', 0.09640831758034027),
('diskutere', 0.08742911153119093),
('anbefaling', 0.06947069943289225),
('delmål', 0.06332703213610585),
('enkelt', 0.05718336483931947),
('paris-aftale', 0.04395085066162571),
('hockeystav', 0.043478260869565216),
('videnskab', 0.04111531190926276),
('diskussion', 0.04064272211720227),
('seriøs', 0.038752362948960305)],
39: [('fossil', 0.08918539325842696),
('bremse', 0.06882022471910113),
('anbefale', 0.061095505617977525),
('total', 0.04845505617977528),
('slippe', 0.04073033707865169),
('risikere', 0.04073033707865169),
('bryde', 0.03089887640449438),
('elektrificering', 0.030196629213483147),
('elektrificere', 0.027387640449438203),
('gasledning', 0.026685393258426966)],
40: [('afgift', 0.39554485249849486),
('skattereform', 0.17278747742323902),
('ensartet', 0.0897049969897652),
('sommerferi', 0.022877784467188442),
('beskatning', 0.02107164358819988),
('modkrav', 0.018663455749548464),
('energiafgift', 0.018061408789885613),
('garanti', 0.018061408789885613),
('provenu', 0.013245033112582781),
('afgifte', 0.012642986152919929)],
41: [('co2', 0.5519828510182208),
('-', 0.4263129689174705),
('neutral', 0.01795284030010718),
('suk', 0.0037513397642015005)],
42: [('transport', 0.15001829491401392),
('køre', 0.0783022319795097),
('kollektiv', 0.05817782656421515),
('infrastruktur', 0.047932674716428834),
('trafik', 0.04463959019392609),
('bus', 0.033662641785583605),
('region', 0.02854006586169045),
('motorvej', 0.028174167581412366),
('tog', 0.019392608854738383),
('udgangspunkt', 0.0190267105744603)],
43: [('virksomhed', 0.07074790643950332),
('mest', 0.07074790643950332),
('sort', 0.05804215997689864),
('betale', 0.04836846664741554),
('lav', 0.04779093271729714),
('betyde', 0.04591394744441236),
('klimaplan', 0.0418712099335836),
('enhedsliste', 0.04042737510828761),
('koste', 0.03190874963904129),
('skat', 0.03190874963904129)],
44: [('bil', 0.3316183348924228),
('elbil', 0.19784845650140317),
('ladestander', 0.0626753975678204),
('fossile', 0.055659494855004675),
('salg', 0.04957904583723106),
('fossilbil', 0.027595884003741813),
('dieselbil', 0.02432179607109448),

('benzinbil', 0.022918615528531337),
 ('bilafgift', 0.018709073900841908),
 ('bilverk', 0.014499532273152479)],
 45: [('amp', 0.06443364868320123),
 ('stige', 0.060598312452058295),
 ('generation', 0.057274354385067755),
 ('lytte', 0.05318332907184863),
 ('lyde', 0.039631807721810276),
 ('ulighed', 0.03912042955765789),
 ('klode', 0.03912042955765789),
 ('ja', 0.03528509332651496),
 ('alvor', 0.03426233699821018),
 ('redde', 0.030427000767067246)],
 46: [('ny', 0.3083097261567517),
 ('år', 0.15372993389990558),
 ('sikre', 0.06817752596789424),
 ('skabe', 0.06628895184135977),
 ('første', 0.06024551463644948),
 ('sikker', 0.052691218130311614),
 ('næste', 0.04268177525967894),
 ('styrke', 0.03947119924457035),
 ('fortsætte', 0.03153918791312559),
 ('bidrage', 0.026817752596789424)],
 47: [('regering', 0.6474321984997115),
 ('ambition', 0.12694748990190421),
 ('politik', 0.11175226005000961),
 ('klimaaftale', 0.033852663973841124),
 ('støtteparti', 0.02308136180034622),
 ('skuffe', 0.019234468166955183),
 ('tempo', 0.01788805539526832),
 ('uambitiøs', 0.010194268128486247),
 ('acceptere', 0.009617234083477592)],
 48: [('krone', 0.21025345622119815),
 ('mio', 0.18778801843317972),
 ('mia', 0.14372119815668202),
 ('milliard', 0.07402073732718895),
 ('million', 0.06394009216589862),
 ('middel', 0.04838709677419355),
 ('afsætte', 0.04349078341013825),
 ('cirka', 0.04003456221198157),
 ('falde', 0.028225806451612902),
 ('halv', 0.018721198156682026)],
 49: [('ramme', 0.15461465271170313),
 ('retfærdig', 0.08468125594671741),
 ('socialt', 0.06517602283539486),
 ('brændstof', 0.05851569933396765),
 ('fly', 0.05042816365366318),
 ('klimaafgift', 0.03615604186489058),
 ('balance', 0.032825880114176975),
 ('check', 0.028544243577545196),
 ('lastbil', 0.02759276879162702),
 ('sverig', 0.021408182683158895)],
 50: [('klar', 0.09860457296570276),
 ('stå', 0.08212844653665098),
 ('parti', 0.06489576328177539),
 ('handling', 0.06363483523873571),
 ('kræve', 0.06027236045729657),
 ('sf', 0.054556153328850034),
 ('folketing', 0.04741089441829186),
 ('plan', 0.04497310020174849),
 ('længe', 0.03597848016139879),

('flertal', 0.0343813046402152)],
 51: [('få', 0.23539850320191344),
 ('godt', 0.12375588303371654),
 ('derfor', 0.09798626649178305),
 ('både', 0.07252526811202839),
 ('lave', 0.05269655119203765),
 ('rigtig', 0.0460612607051925),
 ('danser', 0.04451816989429828),
 ('vigtigt', 0.036262634056014195),
 ('mulighed', 0.03433377054239642),
 ('krav', 0.031941979785510374)],
 52: [('finanslov', 0.3388101983002833),
 ('løft', 0.06345609065155808),
 ('gratis', 0.06288951841359773),
 ('minimumsnormering', 0.04362606232294618),
 ('pædagog', 0.031161473087818695),
 ('psykiatri', 0.029461756373937678),
 ('rødgrøn', 0.020963172804532578),
 ('måles', 0.020396600566572238),
 ('sygeplejerske', 0.01643059490084986),
 ('derude', 0.01643059490084986)],
 53: [('ende', 0.0691515448749387),
 ('hurtig', 0.04659146640510054),
 ('vist', 0.04413928396272683),
 ('slem', 0.0392349190779794),
 ('afskaffe', 0.03776360961255518),
 ('træffe', 0.0358018636586562),
 ('besparelse', 0.03285924472780775),
 ('høring', 0.02550269740068661),
 ('afsløre', 0.02550269740068661),
 ('insistere', 0.02501226091221187)],
 54: [('dyr', 0.15525231719876417),
 ('sted', 0.14572605561277033),
 ('billig', 0.07981462409886715),
 ('pris', 0.05303810504634397),
 ('mindre', 0.04660144181256436),
 ('købe', 0.04042224510813594),
 ('forbrug', 0.03347064881565397),
 ('hurtigt', 0.029351184346035016),
 ('fjerne', 0.028836251287332648),
 ('sænke', 0.027806385169927908)],
 55: [('vej', 0.1382378472222222),
 ('sætte', 0.1330295138888889),
 ('hvordan', 0.0763888888888889),
 ('frem', 0.0693359375),
 ('endnu', 0.06586371527777778),
 ('høj', 0.06108940972222224),
 ('afgørende', 0.04318576388888889),
 ('samtidig', 0.03765190972222224),
 ('klog', 0.033854166666666664),
 ('måde', 0.028645833333333332)],
 56: [('mål', 0.2709477202196228),
 ('reduktion', 0.20148006684172834),
 ('nå', 0.1542134160897589),
 ('klimamål', 0.10503700167104321),
 ('procent', 0.047744091668656),
 ('målsætning', 0.044879446168536646),
 ('pct', 0.03533062783480544),
 ('70', 0.02339460491764144),
 ('målet', 0.013129625208880401),
 ('årlig', 0.011458582000477441)],

57: [('klimakrise', 0.3294964028776978),
 ('løse', 0.14244604316546763),
 ('coronakrise', 0.050359712230215826),
 ('klimaudfordring', 0.050359712230215826),
 ('biodiversitetskrise', 0.036930455635491605),
 ('håndtere', 0.022541966426858514),
 ('information', 0.02158273381294964),
 ('fagbevægelse', 0.017266187050359712),
 ('nøle', 0.015827338129496403),
 ('forside', 0.01486810551558753)],
 58: [('statsminister', 0.12863436123348018),
 ('konkret', 0.11629955947136564),
 ('erhvervsliv', 0.0763582966226138),
 ('frederiksen', 0.0737151248164464),
 ('først', 0.0722466960352423),
 ('finansiering', 0.054038179148311306),
 ('mette', 0.0513950073421439),
 ('klart', 0.050220264317180616),
 ('rejse', 0.03436123348017621),
 ('hverken', 0.030543318649045522)],
 59: [('fri', 0.20604565083281925),
 ('v', 0.08883405305367058),
 ('status', 0.03269586674892042),
 ('tusind', 0.0283775447254781),
 ('vælgererklæring', 0.024059222702035782),
 ('antiracistisk', 0.023442319555829736),
 ('udfordre', 0.023442319555829736),
 ('støjberg', 0.020974706971005553),
 ('løgn', 0.01974090067859346),
 ('bestemme', 0.01850709438618137)],
 60: [('ambitiøs', 0.3210964268232991),
 ('arbejdsplads', 0.2222222222222222),
 ('levere', 0.18404307391091532),
 ('vækst', 0.1429270680372002),
 ('gavne', 0.048458149779735685),
 ('udland', 0.030347528144884974),
 ('dkbiz', 0.01860009789525208),
 ('velstand', 0.016642192853646598),
 ('klimaneutralitet', 0.01566324033284386)],
 61: [('global', 0.13399602385685885),
 ('øge', 0.11848906560636183),
 ('international', 0.068389662027833),
 ('ude', 0.06719681908548708),
 ('parisaftale', 0.04493041749502982),
 ('klimaambition', 0.04254473161033797),
 ('usa', 0.03300198807157058),
 ('forhåbentlig', 0.02345924453280318),
 ('herhjemme', 0.022664015904572565),
 ('præsident', 0.02186878727634195)],
 62: [('bruge', 0.48850174216027875),
 ('penge', 0.2961672473867596),
 ('råderum', 0.04041811846689895),
 ('forhindre', 0.023693379790940768),
 ('halvdel', 0.023344947735191638),
 ('klimaeffekt', 0.021254355400696864),
 ('derudover', 0.020905923344947737),
 ('formål', 0.011498257839721254),
 ('nul', 0.010801393728222997),
 ('afgiftslettelse', 0.010801393728222997)],
 63: [('dkgre', 1.0)],
 64: [(' ', 0.7583337240376952),

('dkpol', 0.23589947958179006),
('endelig', 0.005766796380514792)],
65: [('time', 0.052854122621564484),
('bo', 0.03805496828752643),
('ansætte', 0.0354122621564482),
('ministerium', 0.0354122621564482),
('overhovedet', 0.03171247357293869),
('morten', 0.023255813953488372),
('forholde', 0.022727272727272728),
('stykke', 0.021670190274841437),
('baggrund', 0.021141649048625793),
('person', 0.02061310782241015)],
66: [('tage', 0.2840482308756671),
('holde', 0.11444949594781577),
('hånd', 0.07432298873295118),
('ansvar', 0.06720695789681755),
('problem', 0.06661395532713975),
('deciliter', 0.06582328523423601),
('fortsat', 0.047440205574224154),
('gælde', 0.03518481913421625),
('række', 0.034394149041312515),
('alvorligt', 0.03261514133227911)],
67: [('affald', 0.08214096449390568),
('plastik', 0.0598834128245893),
('genanvendelse', 0.057763645998940114),
('brænde', 0.04663487016428193),
('genbrug', 0.034976152623211444),
('ressource', 0.033386327503974564),
('plast', 0.03126656067832538),
('genanvende', 0.03073661897191309),
('madspild', 0.024377318494965553),
('genbruge', 0.024377318494965553)],
68: [('byggeri', 0.215929203539823),
('trist', 0.10088495575221239),
('amager', 0.08141592920353982),
('fællede', 0.06548672566371681),
('beton', 0.035398230088495575),
('demonstration', 0.03185840707964602),
('klimakrav', 0.03185840707964602),
('fredning', 0.03185840707964602),
('københavn', 0.024778761061946902),
('program', 0.024778761061946902)],
69: [('hav', 0.10325710325710326),
('ødelægge', 0.0990990990990991),
('fiskeri', 0.060291060291060294),
('havmiljø', 0.05197505197505198),
('øresund', 0.04573804573804574),
('havbrug', 0.042966042966042964),
('fisk', 0.030492030492030493),
('havnatur', 0.030492030492030493),
('foregå', 0.029106029106029108),
('forhandlingsbord', 0.026334026334026334)],
70: [('fns', 0.07686832740213523),
('lede', 0.06548042704626335),
('verdensmål', 0.0597864768683274),
('bede', 0.05693950177935943),
('fn', 0.05124555160142349),
('klimatopmøde', 0.03558718861209965),
('åbning', 0.02704626334519573),
('drøftelse', 0.020640569395017794),
('blokere', 0.019217081850533807),

('indlede', 0.018505338078291814)],
 71: [('udledning', 0.2372558361124345),
 ('reducere', 0.19342544068604098),
 ('ton', 0.12577417818008577),
 ('udlede', 0.07908527870414483),
 ('udslip', 0.06955693187232015),
 ('drivhusgasse', 0.04954740352548833),
 ('nedbringe', 0.049070986183897096),
 ('aalborg', 0.04811815150071463),
 ('portland', 0.022391615054787994),
 ('teknologisk', 0.017627441638875654)],
 72: [('igen', 0.4220121539500338),
 ('love', 0.14787305874409182),
 ('stormagt', 0.074949358541526),
 ('socialdemokrati', 0.062120189061444966),
 ('før', 0.04929101958136394),
 ('front', 0.048615800135043886),
 ('guld', 0.02025658338960162),
 ('hedde', 0.019581363943281565),
 ('klimavidenskab', 0.018906144496961513),
 ('sølle', 0.01823092505064146)],
 73: [('naturligt', 0.05860113421550094),
 ('sælge', 0.05103969754253308),
 ('princip', 0.04788909892879647),
 ('forurener', 0.03969754253308128),
 ('klimatosse', 0.03465658475110271),
 ('selskab', 0.03213610586011342),
 ('bære', 0.02835538752362949),
 ('energiø', 0.027725267800882167),
 ('\u200d', 0.02520478890989288),
 ('gigantisk', 0.023944549464398234)],
 74: [('sende', 0.15565729234793982),
 ('finansministerium', 0.052975801177240024),
 ('redskab', 0.03727926749509483),
 ('finansminister', 0.034009156311314584),
 ('nær', 0.027468933943754086),
 ('regnemodel', 0.025506867233485938),
 ('trænge', 0.02419882275997384),
 ('ignorere', 0.02419882275997384),
 ('rød-grøn', 0.02419882275997384),
 ('afgøre', 0.020928711576193592)],
 75: [('dejligt', 0.08818541548897682),
 ('skære', 0.06726964386659129),
 ('landbrugsstøtte', 0.06161673261729791),
 ('tur', 0.054833239118145846),
 ('landbrugsjord', 0.054833239118145846),
 ('turbo', 0.05144149236856981),
 ('udtagning', 0.04918032786885246),
 ('gribe', 0.03335217637083098),
 ('udtage', 0.02882984737139627),
 ('højere', 0.025438100621820236)],
 76: [('hvorfor', 0.16855429713524317),
 ('ord', 0.10526315789473684),
 ('rød', 0.0849433710859427),
 ('blå', 0.07528314457028648),
 ('alene', 0.04563624250499667),
 ('blok', 0.038640906062624915),
 ('borgerlig', 0.03697534976682212),
 ('konservativ', 0.028647568287808126),
 ('nævne', 0.028314457028647568),
 ('kronik', 0.024983344437041973)],

77: [('energieffektivisering', 0.07036423841059603),
('bygning', 0.06456953642384106),
('energirenovering', 0.061258278145695365),
('udenlandsk', 0.057119205298013245),
('skærpe', 0.04470198675496689),
('karensperiode', 0.043874172185430466),
('årig', 0.042218543046357616),
('lejer', 0.04056291390728477),
('styrkelse', 0.0347682119205298),
('overskudsvarme', 0.03145695364238411)],
78: [('barn', 0.23564954682779457),
('uddannelse', 0.15709969788519637),
('skole', 0.06419939577039276),
('elev', 0.035498489425981876),
('lære', 0.027945619335347432),
('folkeskole', 0.0256797583081571),
('voksen', 0.022658610271903322),
('familie', 0.0222809667673716),
('arbejdskraft', 0.01850453172205438),
('bevægelse', 0.01812688821752266)],
79: [('strøm', 0.0938722294654498),
('bygge', 0.09300304215558453),
('vindmøl', 0.08648413733159496),
('havvindmøllepark', 0.07996523250760539),
('havvind', 0.06910039113428944),
('energiaftale', 0.0595393307257714),
('energiforlig', 0.03780964797913951),
('danmarkshistorie', 0.024337244676227728),
('energiøe', 0.02216427640156454),
('kilometer', 0.01912212081703607)],
80: [('kalde', 0.05010211027910143),
('svar', 0.04601769911504425),
('stille', 0.04520081688223281),
('spørgsmål', 0.039618788291354665),
('sag', 0.037848876786929886),
('samråd', 0.03716814159292035),
('minister', 0.03662355343771273),
('passe', 0.031313818924438394),
('svare', 0.025323349217154525),
('tal', 0.02341729067392784)],
81: [('valg', 0.1696522655426765),
('klimaforandring', 0.128907622058307),
('folk', 0.09062170706006323),
('måske', 0.08640674394099052),
('mandat', 0.045310853530031614),
('venstrefløj', 0.035475939585528624),
('demokrati', 0.030909729539866527),
('hold', 0.0273972602739726),
('vælger', 0.026694766420793818),
('maj', 0.018967334035827187)],
82: [('arbejdsmiljø', 0.2026726057906459),
('pension', 0.044543429844097995),
('nedslidt', 0.042316258351893093),
('psykisk', 0.04120267260579064),
('arbejdstilsyn', 0.03841870824053452),
('udsat', 0.03285077951002227),
('syg', 0.02951002227171492),
('pensionsalder', 0.021714922048997772),
('sygehus', 0.021158129175946547),
('dårligt', 0.02060133630289532)],
83: [('kød', 0.11103678929765887),

('spise', 0.09698996655518395),
 ('mad', 0.09096989966555184),
 ('køkken', 0.048160535117056855),
 ('plantebaseret', 0.046153846153846156),
 ('kost', 0.04080267558528428),
 ('kostråd', 0.03946488294314381),
 ('indkøb', 0.03545150501672241),
 ('vegansk', 0.0274247491638796),
 ('måltid', 0.02274247491638796)],
 84: [('♥', 0.22047670639219935),
 ('', 0.1695557963163597),
 ('☺', 0.08829902491874322),
 ('♥', 0.07638136511375948),
 ('☹', 0.06121343445287107),
 ('☹', 0.047670639219934995),
 ('☹', 0.04496208017334778),
 ('♣', 0.03791982665222102),
 ('♣', 0.021668472372697724),
 ('♣', 0.01733477789815818)],
 85: [('offentlig', 0.21454283657307416),
 ('klimavenlig', 0.16846652267818574),
 ('sundhed', 0.1511879049676026),
 ('stat', 0.12526997840172785),
 ('sund', 0.11591072714182865),
 ('gavn', 0.10511159107271419),
 ('bord', 0.06623470122390208),
 ('sikkerhed', 0.05327573794096472)],
 86: [('eu', 0.32456140350877194),
 ('eus', 0.08079409048938135),
 ('europæisk', 0.07663896583564174),
 ('budget', 0.039242843951985226),
 ('fremgang', 0.020313942751615882),
 ('bruxelles', 0.018005540166204988),
 ('kommission', 0.01662049861495845),
 ('europa-parlament', 0.014773776546629732),
 ('energipolitik', 0.014312096029547553),
 ('eu-parlament', 0.013850415512465374)],
 87: [('by', 0.19267955801104972),
 ('miljøminister', 0.16091160220994474),
 ('ren', 0.15469613259668508),
 ('luft', 0.09185082872928177),
 ('dø', 0.04143646408839779),
 ('miljøzon', 0.04005524861878453),
 ('tilgang', 0.03591160220994475),
 ('luftforurening', 0.03383977900552486),
 ('fugl', 0.0324585635359116),
 ('efterlyse', 0.028314917127071824)],
 88: [('dyrevelfærd', 0.18650088809946713),
 ('fødevareminister', 0.037892243931320305),
 ('tvinge', 0.036708111308466546),
 ('gris', 0.03078744819419775),
 ('dyrevelfærde', 0.03019538188277087),
 ('spor', 0.027827116637063352),
 ('lide', 0.027235050325636473),
 ('stramme', 0.02486678507992895),
 ('lovforslag', 0.021906453522794554),
 ('adfærd', 0.020130254588513915)],
 89: [('drikkevand', 0.13692946058091288),
 ('forbud', 0.07745504840940526),
 ('forbyde', 0.0756108805901337),
 ('vand', 0.06685108344859382),

```

('beskyttelse', 0.0437989857076994),
('sprøjtegifte', 0.042876901798063624),
('grundvand', 0.039188566159520516),
('pesticider', 0.039188566159520516),
('sprøjtegift', 0.03503918856615952),
('miljøstyrelse', 0.024896265560165973)],
90: [('skov', 0.3379810373675404),
('urørt', 0.15839375348577803),
('ha', 0.07083100948131622),
('hektar', 0.05800334634690463),
('omlægge', 0.04852203011712214),
('udpege', 0.03067484662576687),
('udlægge', 0.026770775237032907),
('biodiversitetsråd', 0.026213050752928055),
('mangfoldig', 0.016731734523145567),
('naturpark', 0.015616285554935862)],
91: [('træ', 0.12697022767075306),
('biomasse', 0.12084063047285463),
('udfasning', 0.07092819614711034),
('hurtigst', 0.05691768826619965),
('erstatte', 0.04903677758318739),
('cykelsti', 0.040280210157618214),
('kommunal', 0.03940455341506129),
('geotermi', 0.03852889667250438),
('århus', 0.03415061295971979),
('spændt', 0.021891418563922942)],
92: [('bolig', 0.2013129102844639),
('beskæftigelse', 0.11487964989059081),
('alm', 0.10831509846827134),
('renovering', 0.10065645514223195),
('energibesparelse', 0.04923413566739606),
('boligaftale', 0.03391684901531729),
('havvindmøller', 0.03282275711159737),
('beboer', 0.03172866520787746),
('energieffektiv', 0.0262582056892779),
('renovere', 0.02188183807439825)],
93: [('nordsø', 0.15261707988980716),
('olie', 0.1443526170798898),
('oliejagt', 0.10137741046831956),
('gas', 0.06831955922865014),
('udbudsrunde', 0.06611570247933884),
('aflyse', 0.046831955922865015),
('keepitintheground', 0.04022038567493113),
('udfase', 0.03415977961432507),
('tilladelse', 0.023140495867768594),
('slutdato', 0.021487603305785124)],
94: [('produktion', 0.14823756639304683),
('producere', 0.07194591984548528),
('omlægning', 0.04635441815548044),
('animalsk', 0.04007725736359247),
('importere', 0.03959439884113954),
('soja', 0.0386286817962337),
('import', 0.035248672139063256),
('dyrke', 0.03090294543698696),
('svin', 0.027522935779816515),
('halvere', 0.024625784645098986)]]}

```

5. Preparing the data for further analysis

5.1 Making a df of most probable words for each topic

In []:

```
#making a dictionary of topics, words  
dict_topics = model.topics(l=0,n=10)
```

```
/content/hSBM_Topicmodel/sbmtm.py:634: RuntimeWarning: invalid value encountered in true_divide  
    p_td_d = (n_db/np.sum(n_db,axis=1)[: ,np.newaxis]).T  
/content/hSBM_Topicmodel/sbmtm.py:640: RuntimeWarning: invalid value encountered in true_divide  
    p_tw_d = (n_dbw/np.sum(n_dbw,axis=1)[: ,np.newaxis]).T
```

In []:

```
my_dict = dict_topics
```

In []:

```
df_test = pd.DataFrame(list(my_dict.items()),columns = ['column1','column2'])
```

In []:

```
#Converting the dict into a df:  
#this gives us a df of the 10 most probable words for each topic  
import numpy as np  
for i in range(len(my_dict)):  
    if i == 0:  
        v = np.hstack((i*np.ones((len(my_dict[i]),1)).astype(int), np.array(my_dict[i])))  
        continue  
    h = np.hstack((i*np.ones((len(my_dict[i]),1)).astype(int), np.array(my_dict[i])))  
    v = np.vstack((v,h))  
new_topic = pd.DataFrame(v)
```

In []:

```
#Renaming the columns:  
dictwords_df = new_topic.rename(columns={0: 'Topic', 1: 'Word', 2: 'Proportion'})
```

In []:

```
dictwords_df
```

Out[]:

	Topic	Word	Proportion
0	0	danmark	0.9349667271627344
1	0	forrest	0.034482758620689655
2	0	førertrøje	0.018753781004234724
3	0	her	0.011796733212341199
4	1	mange	0.27047712099006677
...
903	94	soja	0.0386286817962337
904	94	import	0.035248672139063256
905	94	dyrke	0.03090294543698696
906	94	svin	0.027522935779816515
907	94	halvere	0.024625784645098986

908 rows × 3 columns

In []:

```
#saving the df as a csv file
from google.colab import files

dictwords_df.to_csv('dictwords_df.csv')
files.download('dictwords_df.csv')
```

5.2 Making a df of probability of document belonging to topic=i

In []:

```
#This gives us a dictionary of the different parameters (there are 6) of the model (in
numpy arrays) - we are interested in p_tw_d
#p_tw_d is an array of the probabilities of a given document (in this case tweet) belong
ing to the different topics
#remember to define the level
model_params = model.get_groups(l=0)
```

```
/content/hSBM_Topicmodel/sbmtm.py:634: RuntimeWarning: invalid value encountered in true_divide
```

```
    p_td_d = (n_db/np.sum(n_db,axis=1)[: ,np.newaxis]).T
```

```
/content/hSBM_Topicmodel/sbmtm.py:640: RuntimeWarning: invalid value encountered in true_divide
```

```
    p_tw_d = (n_dbw/np.sum(n_dbw,axis=1)[: ,np.newaxis]).T
```

In []:

```
#we see that the shape corresponds to (topics, documents)  
model_params['p_tw_d'].shape
```

Out[]:

(95, 26052)

In []:

```
doc_prob = model_params['p_tw_d']
```

In []:

```
doc_prob[0].shape
```

Out[]:

(26052,)

In []:

```
#resetting the index, so there is no gaps. This makes it easier to concat the two dataframes properly.  
df_new = df.reset_index()
```

In []:

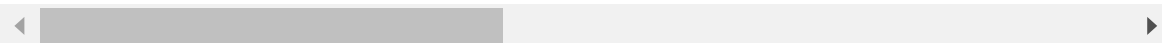
```
df_new
```

Out[]:

	index	Unnamed: 0	Unnamed: 0.1	Unnamed: 0.1.1	Unnamed: 0.1.1.1	id	created_at
0	0	0	5068	7155	7155.0	1387387662387105793	2021-04-28 12:46:09
1	1	2	5083	7170	7170.0	1380080544990519297	2021-04-08 08:50:17
2	2	3	5090	7177	7177.0	1333500247021146116	2020-11-30 19:56:48
3	3	4	5122	7209	7209.0	1098702372417798146	2019-02-21 21:53:53
4	4	5	5133	7220	7220.0	1055881691225604097	2018-10-26 17:59:47
...
26047	26047	31237	311258	311261	NaN	980488133107224582	2018-04-01 16:52:42
26048	26048	31238	311301	311304	NaN	966366728707104768	2018-02-21 17:39:17
26049	26049	31239	311310	311313	NaN	964434451852906496	2018-02-16 09:41:06

	index	Unnamed: 0	Unnamed: 0.1	Unnamed: 0.1.1	Unnamed: 0.1.1.1	id	created_at
26050	26050	31240	311329	311332	NaN	959966153497108485	2018-02-04 01:45:40
26051	26051	31242	311378	311381	NaN	948471815462146048	2018-01-03 08:31:17

26052 rows × 14 columns



In []:

```
#making the array into a dataframe
doc_prob = pd.DataFrame(doc_prob)
```

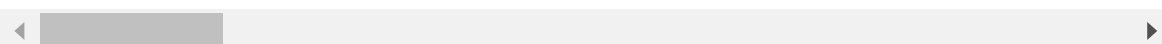
In []:

```
#transposing the dataframe, rows become columns, columns become rows
doc_prob1 = doc_prob.transpose()
doc_prob1
```

Out[]:

	0	1	2	3	4	5	6	7	
0	0.111111	0.111111	0.111111	0.111111	0.111111	0.111111	0.111111	0.111111	0.111111
1	0.000000	0.000000	0.000000	0.000000	0.222222	0.000000	0.111111	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.166667	0.166667
3	0.000000	0.000000	0.000000	0.000000	0.200000	0.000000	0.200000	0.000000	0.400000
4	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
...
26047	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.090909	0.000000
26048	0.000000	0.000000	0.000000	0.066667	0.000000	0.000000	0.000000	0.066667	0.000000
26049	0.000000	0.000000	0.000000	0.083333	0.000000	0.000000	0.000000	0.166667	0.000000
26050	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.052632	0.000000
26051	0.000000	0.100000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

26052 rows × 95 columns



In []:

```
#concatenating the original dataframe with the doc probabilities dataframe
hsbm_model = pd.concat([df_new, doc_prob1], axis=1)
```


In []:

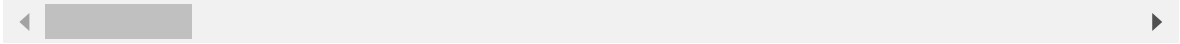
```
#this gives us a new dataframe  
#each row is a document  
#each document has a probability of belonging to topics (in the columns)  
hsbm_model
```

Out[]:

	index	Unnamed: 0	Unnamed: 0.1	Unnamed: 0.1.1	Unnamed: 0.1.1.1	id	created_at
0	0	0	5068	7155	7155.0	1387387662387105793	2021-04-28 12:46:09
1	1	2	5083	7170	7170.0	1380080544990519297	2021-04-08 08:50:17
2	2	3	5090	7177	7177.0	1333500247021146116	2020-11-30 19:56:48
3	3	4	5122	7209	7209.0	1098702372417798146	2019-02-21 21:53:53
4	4	5	5133	7220	7220.0	1055881691225604097	2018-10-26 17:59:47
...
26047	26047	31237	311258	311261	NaN	980488133107224582	2018-04-01 16:52:42
26048	26048	31238	311301	311304	NaN	966366728707104768	2018-02-21 17:39:17
26049	26049	31239	311310	311313	NaN	964434451852906496	2018-02-16 09:41:06

	index	Unnamed: 0	Unnamed: 0.1	Unnamed: 0.1.1	Unnamed: 0.1.1.1	id	created_at
26050	26050	31240	311329	311332	NaN	959966153497108485	2018-02-04 01:45:40
26051	26051	31242	311378	311381	NaN	948471815462146048	2018-01-03 08:31:17

26052 rows × 109 columns



Denmark as frontrunner

```
{0: [('danmark', 0.9349667271627344),  
      ('forrest', 0.034482758620689655),  
      ('førertrøje', 0.018753781004234724),  
      ('her', 0.011796733212341199)],
```

Agriculture

```
4: [('landbrug', 0.45383324006715164),  
     ('økologi', 0.07330721880246223),  
     ('økologisk', 0.06603245663122552),  
     ('fødevare', 0.06323447118074986),  
     ('landmand', 0.04280917739227756),  
     ('plantebasere', 0.03273642977056519),  
     ('landbrugspakke', 0.01846670397313934),  
     ('kø', 0.01650811415780638),  
     ('mark', 0.013989927252378288),  
     ('fordoble', 0.012870733072188025)],
```

Green energy

```
24: [('energi', 0.3205013428827216),  
      ('foregangsland', 0.07699194270367055),  
      ('vedvarende', 0.07341092211280215),  
      ('energieffektivitet', 0.04028648164726947),  
      ('energiudspil', 0.03849597135183527),  
      ('vind', 0.03715308863025962),  
      ('fastholde', 0.03581020590868397),  
      ('dække', 0.02954341987466428),  
      ('sol', 0.026410026857654433),  
      ('elafgift', 0.021933751119068933)],
```

Research and innovation

```
29: [('forskning', 0.2718064153066967),  
      ('ekstra', 0.10635903207653348),  
      ('massivt', 0.04952166572875633),  
      ('bedst', 0.03432751828925155),  
      ('innovation', 0.032639279684862126),  
      ('satsning', 0.032076533483398985),  
      ('holland', 0.023635340461451885),  
      ('tilfreds', 0.023635340461451885),  
      ('dertil', 0.020258863252673044),  
      ('pulje', 0.019696117051209903)],
```

Nature conservation

```
30: [('natur', 0.3303988261971455),  
      ('biodiversitet', 0.07362945178071229),  
      ('plads', 0.060157396291850074),  
      ('vild', 0.04668534080298786),  
      ('beskytte', 0.029611844737895156),  
      ('naturnationalpark', 0.026277177537681738),  
      ('dknatur', 0.024409763905562223),  
      ('areal', 0.023209283713485393),  
      ('true', 0.022408963585434174),  
      ('plante', 0.01854074963318661)],
```

Economy

```
33: [('økonomi', 0.16730917501927525),  
      ('økonomisk', 0.1661526599845798),  
      ('krise', 0.09020817270624518),  
      ('ansvarlig', 0.06399383191981496),  
      ('genstart', 0.0551272166538165),  
      ('genopretning', 0.04934464148033924),
```

('langsigtet', 0.028141865844255976),
('solidarisk', 0.027756360832690823),
('hjælpepak', 0.020046260601387818),
('kickstartet', 0.016962220508866616)],

Climate law

35: [('klimalov', 0.33577586206896554),
('binde', 0.1254310344827586),
('vedtage', 0.07887931034482759),
('borgerforslag', 0.06853448275862069),
('reduktionsmål', 0.04008620689655172),
('forpligte', 0.03405172413793103),
('opposition', 0.03146551724137931),
('folketingsal', 0.027155172413793102),
('jul', 0.01939655172413793),
('ommer', 0.014224137931034483)],

Transportation

42: [('transport', 0.15001829491401392),
('køre', 0.0783022319795097),
('kollektiv', 0.05817782656421515),
('infrastruktur', 0.047932674716428834),
('trafik', 0.04463959019392609),
('bus', 0.033662641785583605),
('region', 0.02854006586169045),
('motorvej', 0.028174167581412366),
('tog', 0.019392608854738383),
('udgangspunkt', 0.0190267105744603)],

Vehicles

44: [('bil', 0.3316183348924228),
('elbil', 0.19784845650140317),
('ladestander', 0.0626753975678204),
('fossile', 0.055659494855004675),
('salg', 0.04957904583723106),
('fossilbil', 0.027595884003741813),
('dieselbil', 0.02432179607109448),
('benzinbil', 0.022918615528531337),
('bilafgift', 0.018709073900841908),
('bilpark', 0.014499532273152479)],

Investment

48: [('krone', 0.21025345622119815),
('mio', 0.18778801843317972),
('mia', 0.14372119815668202),
('milliard', 0.07402073732718895),
('million', 0.06394009216589862),
('middel', 0.04838709677419355),
('afsætte', 0.04349078341013825),
('cirka', 0.04003456221198157),
('falde', 0.028225806451612902),
('halv', 0.018721198156682026)],

Reaching the climate goal

56: [('mål', 0.2709477202196228),
('reduktion', 0.20148006684172834),
('nå', 0.1542134160897589),
('klimamål', 0.10503700167104321),
('procent', 0.047744091668656),
('målsætning', 0.044879446168536646),
('pct', 0.03533062783480544),

```
('70', 0.02339460491764144),
('målet', 0.013129625208880401),
('årlig', 0.011458582000477441)],
```

Climate crisis

```
57: [('klimakrise', 0.3294964028776978),
('løse', 0.14244604316546763),
('coronakrise', 0.050359712230215826),
('klimaudfordring', 0.050359712230215826),
('biodiversitetskrise', 0.036930455635491605),
('håndtere', 0.022541966426858514),
('information', 0.02158273381294964),
('fagbevægelse', 0.017266187050359712),
('nøle', 0.015827338129496403),
('forside', 0.01486810551558753)],
```

Denmark in the world

```
61: [('global', 0.13399602385685885),
('øge', 0.11848906560636183),
('international', 0.068389662027833),
('ude', 0.06719681908548708),
('parisaftale', 0.04493041749502982),
('klimaambition', 0.04254473161033797),
('usa', 0.03300198807157058),
('forhåbentlig', 0.02345924453280318),
('herhjemme', 0.022664015904572565),
('præsident', 0.02186878727634195)],
```

Reuse/recycle

```
67: [('affald', 0.08214096449390568),
('plastik', 0.0598834128245893),
('genanvendelse', 0.057763645998940114),
('brænde', 0.04663487016428193),
('genbrug', 0.034976152623211444),
('ressource', 0.033386327503974564),
('plast', 0.03126656067832538),
('genanvende', 0.03073661897191309),
('madspild', 0.024377318494965553),
('genbruge', 0.024377318494965553)],
```

Ocean

```
69: [('hav', 0.10325710325710326),
('ødelægge', 0.0990990990990991),
('fiskeri', 0.060291060291060294),
('havmiljø', 0.05197505197505198),
('øresund', 0.04573804573804574),
('havbrug', 0.042966042966042964),
('fisk', 0.030492030492030493),
('havnatur', 0.030492030492030493),
('foregå', 0.029106029106029108),
('forhandlingsbord', 0.026334026334026334)],
```

FN

```
70: [('fns', 0.07686832740213523),
('lede', 0.06548042704626335),
('verdensmål', 0.0597864768683274),
('bede', 0.05693950177935943),
('fn', 0.05124555160142349),
('klimatopmøde', 0.03558718861209965),
('åbning', 0.02704626334519573),
('drøftelse', 0.020640569395017794),
```

('blokere', 0.019217081850533807),
('indlede', 0.018505338078291814)],

Emissions

71: [('udledning', 0.2372558361124345),
('reducere', 0.19342544068604098),
('ton', 0.12577417818008577),
('udlede', 0.07908527870414483),
('udslib', 0.06955693187232015),
('drivhusgasse', 0.04954740352548833),
('nedbringe', 0.049070986183897096),
('aalborg', 0.04811815150071463),
('portland', 0.022391615054787994),
('teknologisk', 0.017627441638875654)],

Windmills

79: [('strøm', 0.0938722294654498),
('bygge', 0.09300304215558453),
('vindmøl', 0.08648413733159496),
('havvindmøllepark', 0.07996523250760539),
('havvind', 0.06910039113428944),
('energi aftale', 0.0595393307257714),
('energi forlig', 0.03780964797913951),
('danmarkshistorie', 0.024337244676227728),
('energi øe', 0.02216427640156454),
('kilometer', 0.01912212081703607)],

Food

83: [('kød', 0.11103678929765887),
('spise', 0.09698996655518395),
('mad', 0.09096989966555184),
('køkken', 0.048160535117056855),
('plantebaseret', 0.046153846153846156),
('kost', 0.04080267558528428),
('kostråd', 0.03946488294314381),
('indkøb', 0.03545150501672241),
('vegansk', 0.0274247491638796),
('måltid', 0.02274247491638796)],

European cooperation

86: [('eu', 0.32456140350877194),
('eus', 0.08079409048938135),
('europæisk', 0.07663896583564174),
('budget', 0.039242843951985226),
('fremgang', 0.020313942751615882),
('bruxelles', 0.018005540166204988),
('kommission', 0.01662049861495845),
('europa-parlament', 0.014773776546629732),
('energi politik', 0.014312096029547553),
('eu-parlament', 0.013850415512465374)],

Water

89: [('drikkevand', 0.13692946058091288),
('forbud', 0.07745504840940526),
('forbyde', 0.0756108805901337),
('vand', 0.06685108344859382),
('beskyttelse', 0.0437989857076994),
('sprøjtegifte', 0.042876901798063624),
('grundvand', 0.039188566159520516),
('pesticider', 0.039188566159520516),
('sprøjtegift', 0.03503918856615952),

('miljøstyrelse', 0.024896265560165973)],

Reforestation

90: [('skov', 0.3379810373675404),
('urørt', 0.15839375348577803),
('ha', 0.07083100948131622),
('hektar', 0.05800334634690463),
('omlægge', 0.04852203011712214),
('udpege', 0.03067484662576687),
('udlægge', 0.026770775237032907),
('biodiversitetsråd', 0.026213050752928055),
('mangfoldig', 0.016731734523145567),
('naturpark', 0.015616285554935862)],

Dirty energy

93: [('nordsø', 0.15261707988980716),
('olie', 0.1443526170798898),
('oliejagt', 0.10137741046831956),
('gas', 0.06831955922865014),
('udbudsrunde', 0.06611570247933884),
('aflyse', 0.046831955922865015),
('keepitintheground', 0.04022038567493113),
('udfase', 0.03415977961432507),
('tilladelse', 0.023140495867768594),
('slutdato', 0.021487603305785124)],

Party	Denmark as frontrunner	Agriculture	Clean energy	Research and innovation	Nature conservation	Economy	Climate law	Transportation	Vehicles	Investment
ALT	0,00538373	0,008527024	0,003053316	0,003135136	0,02178595	0,004075329	0,007524479	0,009527827	0,003877459	0,007294044
DF	0,007576168	0,020178976	0,003260782	0,002949107	0,013370161	0,002827377	0,001978239	0,006109099	0,002163748	0,004934316
EL	0,005856573	0,012875071	0,0044222	0,002558951	0,015149237	0,005835941	0,006667858	0,006262063	0,004056055	0,009086814
KF	0,009669958	0,00459807	0,010442903	0,003685612	0,015252914	0,003548801	0,009764238	0,003691413	0,004934892	0,007502629
LA	0,004325615	0,010911581	0,009946338	0,00850584	0,013058008	0,006700043	0,003176237	0,009647551	0,013425423	0,012509587
NB	0,011906936	0,028622435	0	0,001547988	0,009325137	0,003998968	0,010550627	0	0	0,009580203
RV	0,010803697	0,00546093	0,00347882	0,004067917	0,006282855	0,010709651	0,004351668	0,004267686	0,003760345	0,005971215
S	0,010324185	0,008582708	0,006212632	0,00628401	0,018724952	0,005457632	0,00532905	0,007469344	0,00486923	0,008817983
SF	0,005608209	0,007061678	0,003803232	0,002258333	0,01681208	0,006768017	0,007766235	0,00814984	0,005166965	0,004974326
V	0,010816506	0,007243105	0,0088659	0,004705914	0,01137508	0,006252425	0,004300698	0,008777094	0,007302671	0,009730838

Party	Reaching the climate goal	Climate crisis	Denmark in the world	Reuse/re cycle	Ocean	Emissions	Windmills	Food	European community	Animal welfare	Water	Reforestation	Dirty energy
ALT	0,008926	0,008453	0,005337	0,002555	0,003003	0,004822	0,002632	0,00511	0,002298	0,003312	0,005779	0,004857	0,01172
DF	0,006824	0,00435	0,004762	0,002077	0,007575	0,004189	0,003053	0,006588	0,004658	0,011595	0,003648	0,00275	0,003927
EL	0,009904	0,004427	0,007069	0,002777	0,006393	0,00571	0,003472	0,002439	0,005873	0,006252	0,01011	0,003811	0,004619
KF	0,00883	0,004411	0,005792	0,007797	0,001838	0,002214	0,007101	0,002484	0,005227	0,001711	0,003609	0,006653	0,001083
LA	0,007552	0,01028	0,003886	0,00159	0,001457	0,003684	0,003908	0	0,006151	0,001019	0,002086	0,005645	0,001027
NB	0,001634	0,006237	0	0	0	0,018596	0	0,003096	0,006976	0,011801	0,006046	0	0,001961
RV	0,011077	0,005501	0,005577	0,001705	0,00146	0,003494	0,003785	0,002939	0,004321	0,001569	0,001826	0,001166	0,001302
S	0,00842	0,003247	0,006229	0,006062	0,002406	0,004458	0,008847	0,003896	0,004568	0,004027	0,005732	0,003478	0,002562
SF	0,011832	0,00698	0,005654	0,004091	0,002509	0,003523	0,003089	0,002999	0,006158	0,003075	0,004795	0,004324	0,008383
V	0,006275	0,003062	0,006646	0,007482	0,001585	0,004452	0,007852	0,001835	0,007086	0,00457	0,003407	0,006002	0,001813