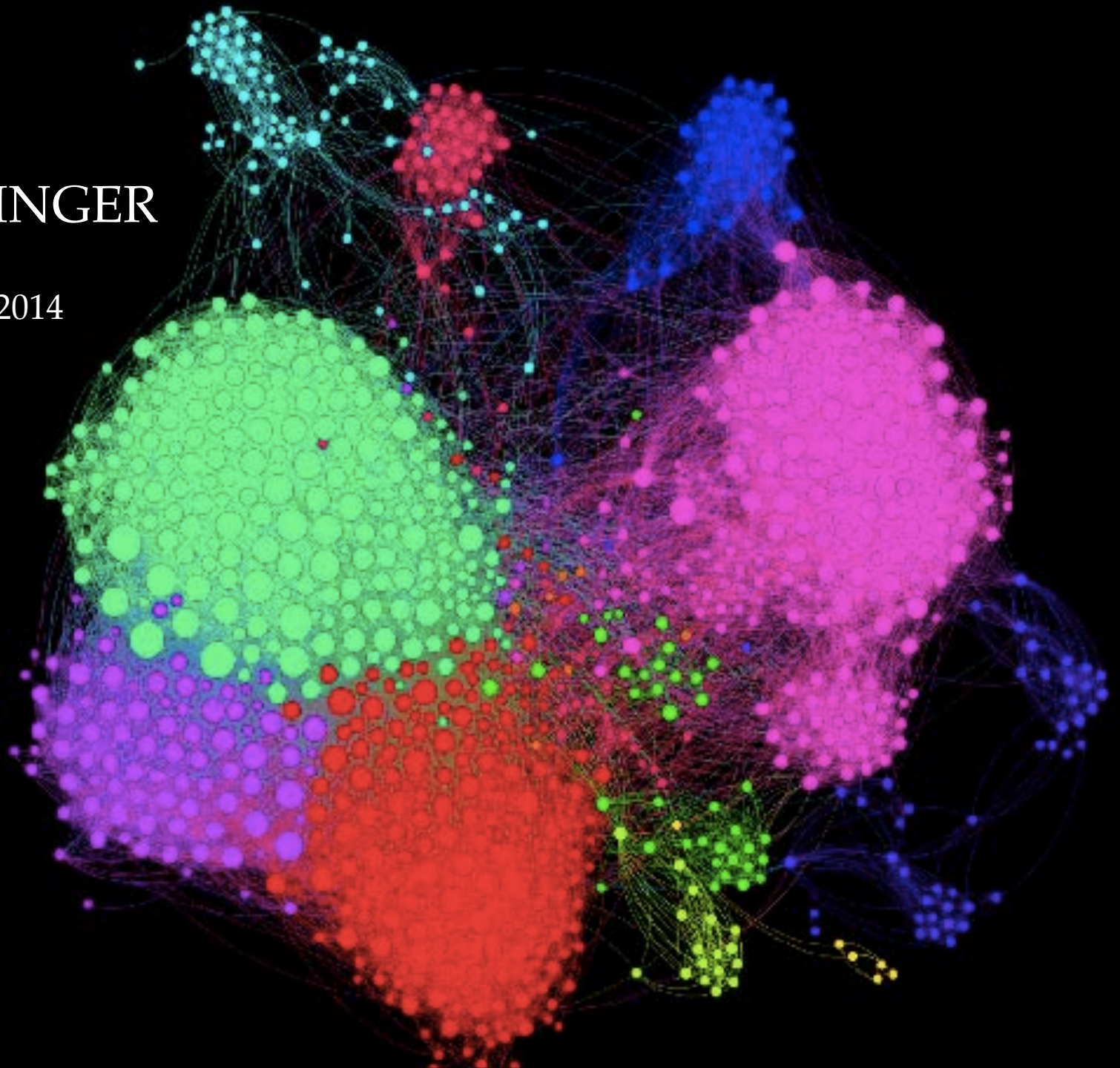


BAG OM BIG DATA VISUALISERINGER

Forskningens Døgn 2014



Mine Tilhørsforhold



BIG DATA FORUM

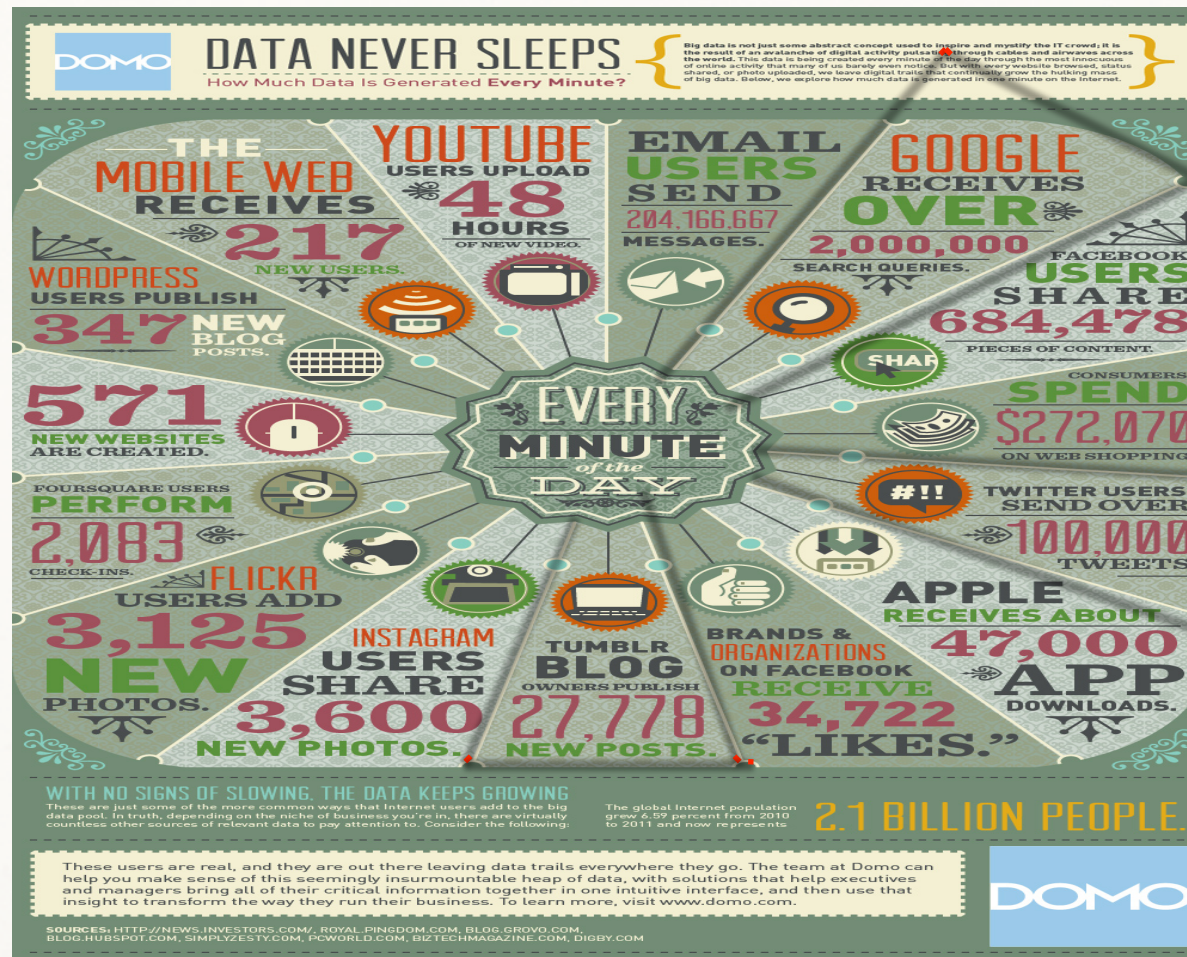


De næste 20 minutter

- ❖ Kort intro til web-baseret Big Data
 - ❖ 5 momenter i produktionen af web-baserede Big Data visualiseringer
 - ❖ FN og organisatoriske dilemmaer i forhold til brugen af web-baseret Big Data
-

Web-Baseret Big Data

Knytter sig til eksplosionen af digitalt data på nettet



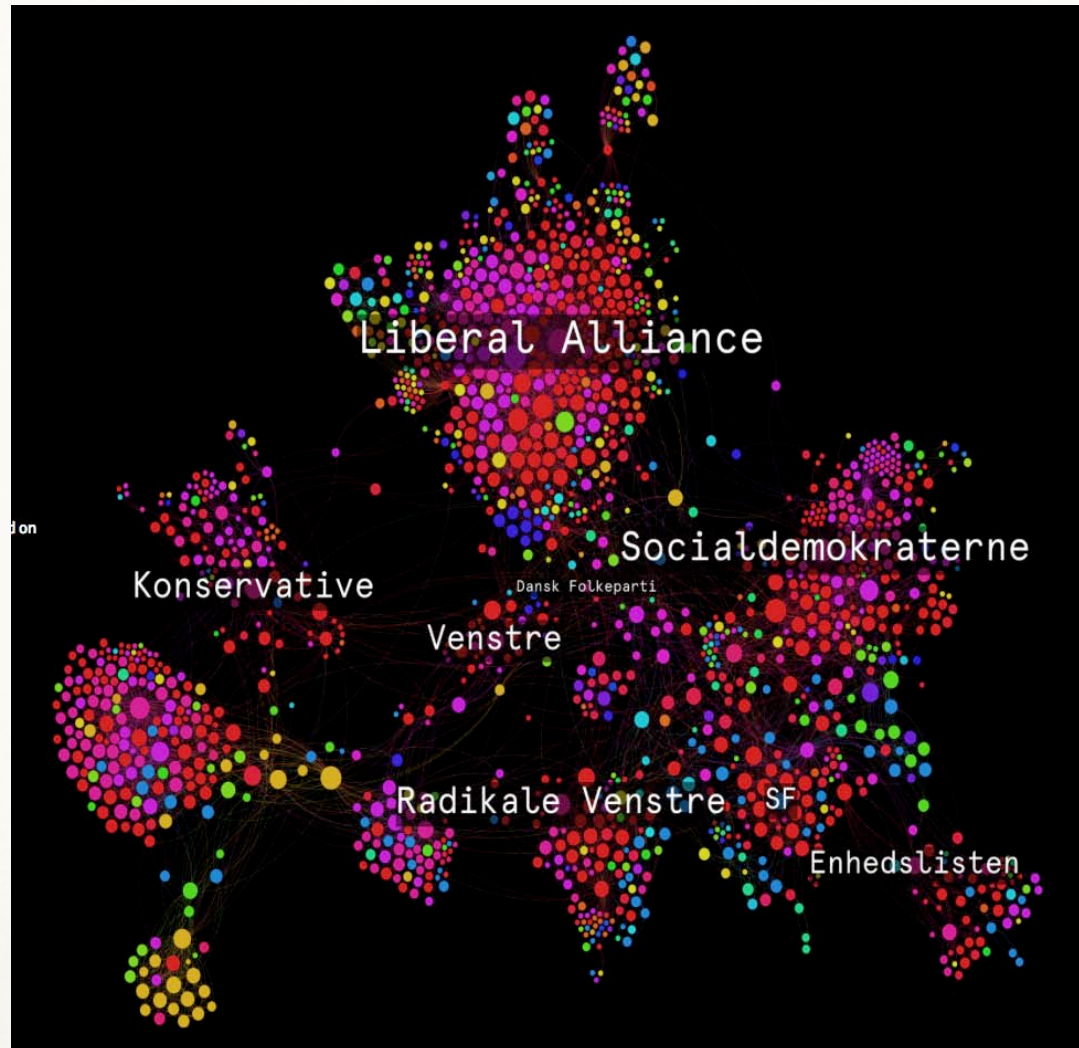
Løs Definition

Web-baseret Big Data består i at indsamle digitale spor fra nettet og (gen)bruge dem som fundament for visualiseringer, der kan fungere som navigationsværktøjer i en kompleks verden.

Google som standard reference



CADM



5 Momenter

i produktionen af web-baserede Big Data visualiseringer

5 Momenter

	1	2	3	4	5
Formål	Produktion	Lagring	Organisering	Distribution	Analyse og visualisering
Centrale Teknologier					

1. Moment - Data Produktion

Hardware



Software Interfaces

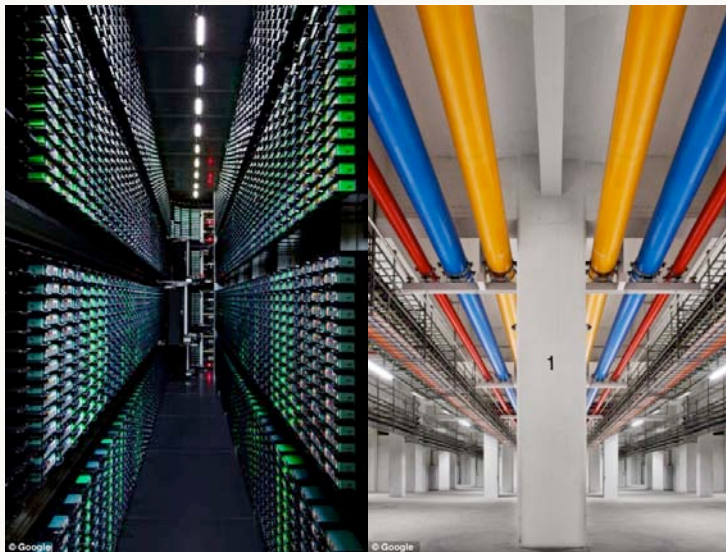


EULAer

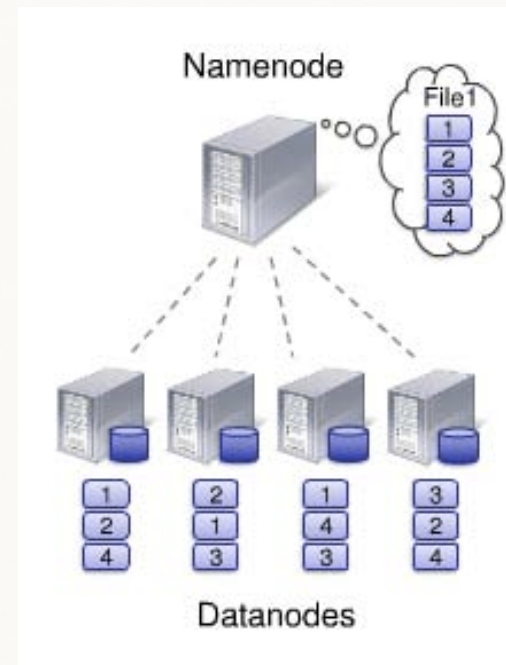


2. Moment - Lagring

Server Farme



Hadoop



3. Moment - Organisering



Et alternativ til relationelle databaser og SQL

3. Moment - Organisering

Relationelle databaser og SQL

The diagram illustrates a table with the following data:

Student ID	Name	Age	Sex	Stress Level
65646	Klaus	20	M	45
93847	Birgitte	18	F	35
63526	Asger	22	M	12
3947	Trine	27	F	67

Annotations:

- Entities:** Student ID
- Attributes:** Name, Age, Sex, Stress Level
- Constraint 1: Cannot be changed:** Name
- Constraint 2: Mutually exclusive values:** Sex

3. Moment - Organisering

Not
Only SQL

Et alternativ til relationelle databaser og SQL

```
{  
  FirstName: "Bob",  
  Address: "5 Oak St.",  
  Hobby: "sailing"  
}
```

```
{  
  FirstName: "Jonathan",  
  Address: "15 Wanamassa Point Road",  
  Children: [  
    {Name: "Michael", Age: 10},  
    {Name: "Jennifer", Age: 8},  
    {Name: "Samantha", Age: 5},  
    {Name: "Elena", Age: 2}  
  ]  
}
```

4. Moment - Distribution

APIer



4. Moment - Distribution

Facebooks Graph API

A Facebook **Group**. The **User**, **Page** and **Application** objects have **groups** connections.

To read a **Group**, you need:

- any valid **access_token** if the group is public (i.e. the group's privacy setting is **OPEN**)
- **user_groups** permission for a user's non-public groups
- **friends_groups** permission for a user's friend's non-public groups
- Groups for Apps and Games require the use of an App **access_token**.

Example

The Facebook Developers **Group**:


<https://graph.facebook.com/195466193802264>

Fields

Name	Description	Permissions	Returns
cover	The URL for the group's cover photo	generic access_token , user_groups , or friends_groups	array containing a valid URL, cover_id and image offset
description	A brief description of the group	generic access_token , user_groups , or friends_groups	string

4. Moment - Distribution

Netvizz



netvizz

[Go to App](#) [Cancel](#)

1,000 people use this app

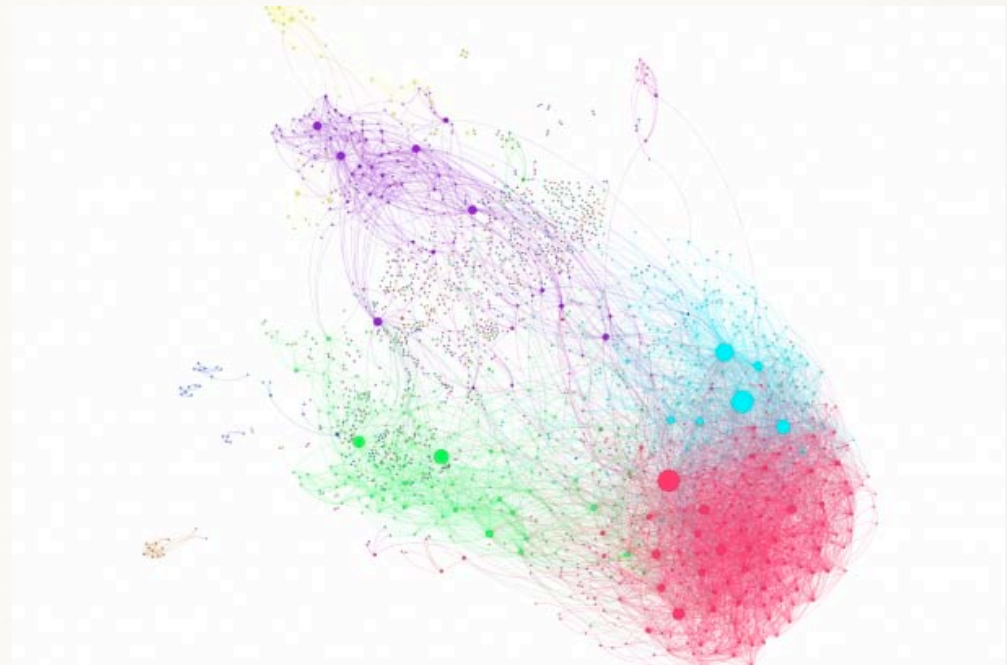
ABOUT THIS APP
Provides data in standard formats (graph and tabular) for your personal network, groups you are a member of, and pages you liked.

Who can see posts this app makes for you on your Facebook timeline: [?]
[Friends](#)

By proceeding, you agree to the netvizz [Privacy Policy](#) - [Report App](#)

THIS APP WILL RECEIVE:

- Your basic info [?]
- Your groups
- Your likes
- Your status updates
- Friends' likes



5. Moment - Analyse

Algoritmer

“Et sæt af logisk definerede instruktioner, der tilsammen angiver, hvordan et specifikt problem løses”

5. Moment - Analyse

Eksempel på simpel algoritme til kortsortering

- a) Tag et tilfældigt kort fra bunken
- b) Gå bunken igennem - alle dem som er højere end dit tilfældige kort, lægger du i en bunke til venstre og alle der er lavere i en bunke til højre.
- c) Læg dit tilfældige kort i venstre bunke
- d) Hvis der er mere end to kort i hver bunke, så gentag pkt 1-3 for hver bunke.
- e) Hvis der er to kort i bunken, så læg det højeste øverst og læg bunken ovenpå den, den blev skilt fra.

5. Moment - Analyse

Google's PageRank algoritme

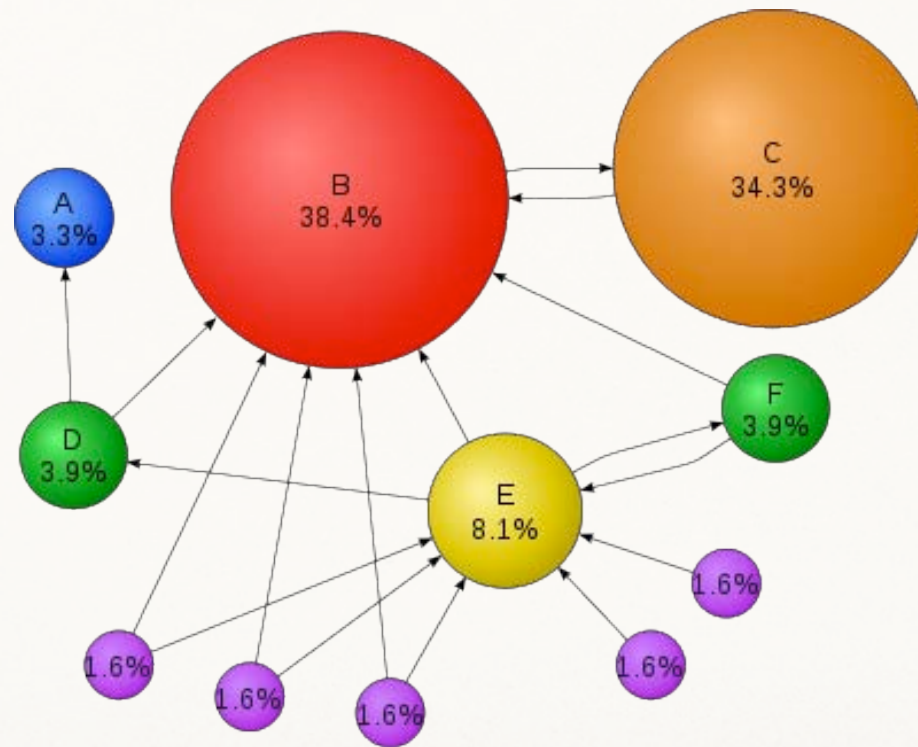
$$\text{PageRank of site} = \sum \frac{\text{PageRank of inbound link}}{\text{Number of links on that page}}$$

OR

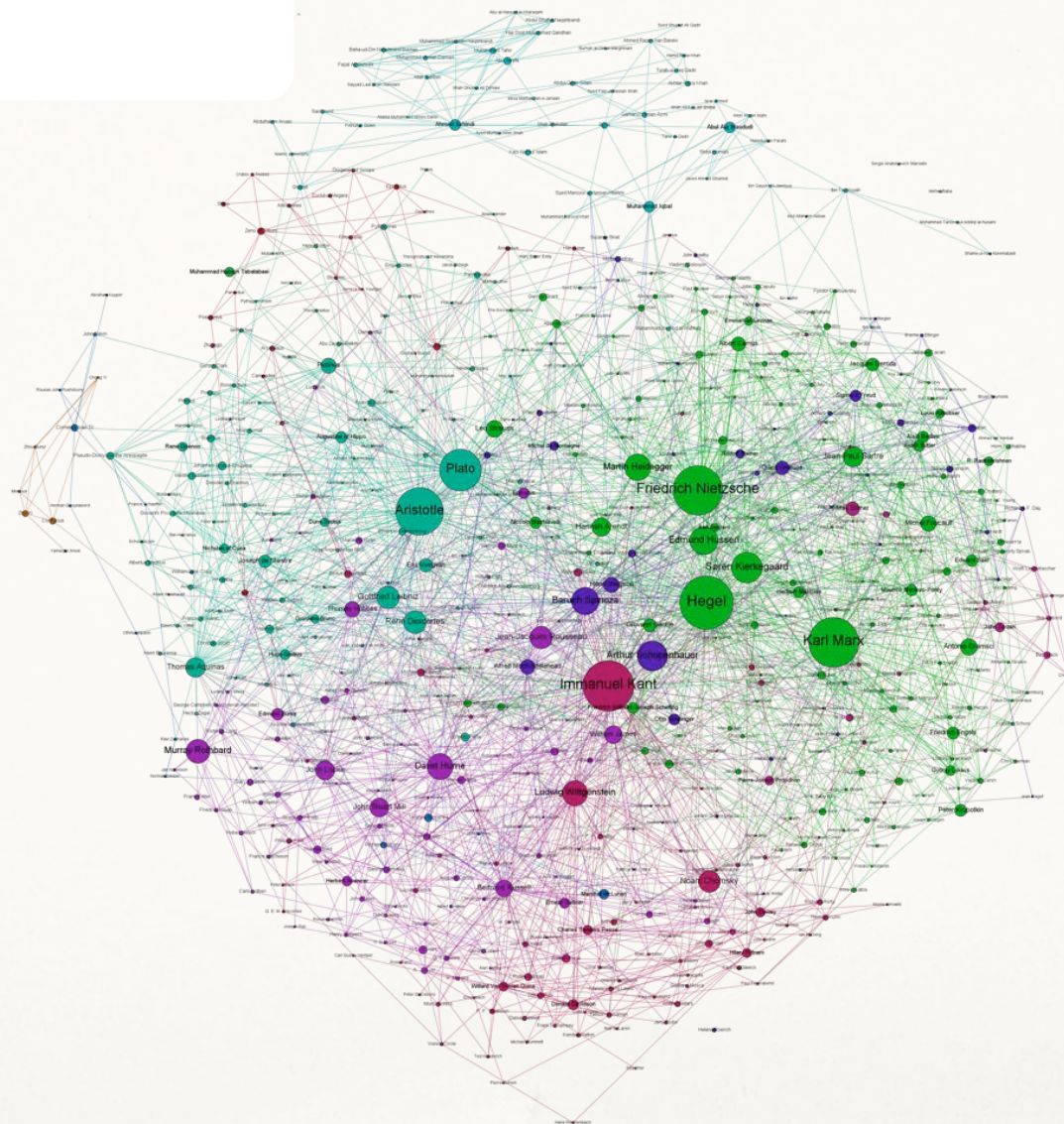
$$PR(u) = (1 - d) + d \times \sum \frac{PR(v)}{N(v)}$$

5. Moment - Analyse

Google's PageRank algoritme



Outputtet af de 5 momenter er visualiseringer

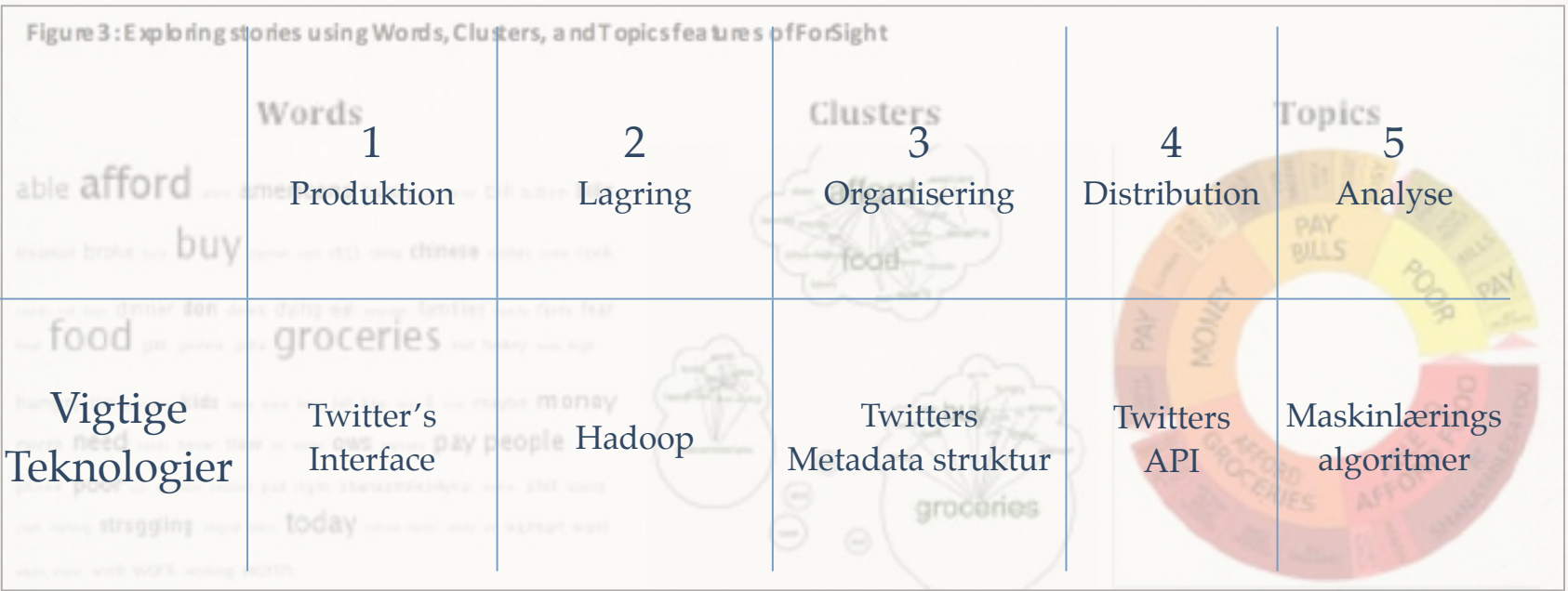


FN & Organisatoriske Dilemmaer

FNs visualisering af krise-signaler

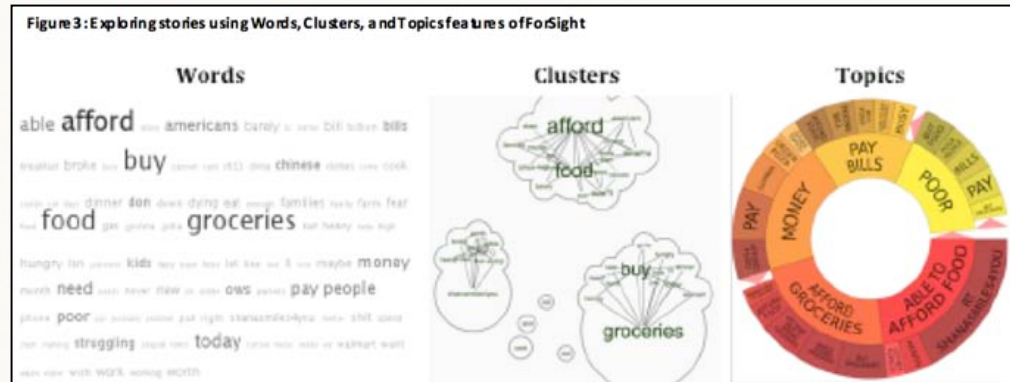
Figure 3: Exploring stories using Words, Clusters, and Topics features of FoSight





Momenter, der er udenfor FNs kontrol

Kilde til organisatoriske dilemmaer i FN



Tweets er pre-strukturerede, når de kommer fra Twitters API

140 tegn

Bruger genereret meta-data (#, @ etc.)

Lav transparens

Konstant forandring

Analysen af det semantiske data skal bygge på automatiserede algoritmer

Induktiv mønstergenkendelse

vs

Organisatorisk resonans

Brugen af web-baseret Big Data skaber nogle organisatoriske og epistemologiske dilemmaer, der nødvendiggør at vi gentænker den måde vi skaber viden om samfundet på.

Diskussioner omkring sådan en gentænkning drejer sig om de fem momenter og de finder sted både i bureaukratiske organisationer som FN og i forskningsgrupper som CADM.

|

Kontakt:

Anders Koed Madsen
akma@learning.aau.dk



forskningsgruppen for
tekno-antropologi