

# **LAPORAN TUGAS EKSPLORASI**

Diajukan untuk memenuhi tugas mata kuliah IF4041 Teknologi Basis Data

oleh :

Ramos Janoah Hasudungan 13514089



**PROGRAM STUDI TEKNIK INFORMATIKA  
SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA  
INSTITUT TEKNOLOGI BANDUNG  
2017**

# 1. Eksplorasi Visualisasi Data

Pada eksplorasi visualisasi data ini, mencakup hal-hal sebagai berikut:

1. Eksplorasi Tools

Tools yang dieksplorasi adalah *gephi*. Gephi adalah salah satu perangkat lunak yang khusus dibuat untuk melakukan visualisasi terhadap data.

2. Pembuatan Visualisasi Menggunakan Gephi

Setelah eksplorasi tools, dilakukan juga pembuatan visualisasi graph yaitu data keterhubungan antar tokoh *game of thrones* dan FB.

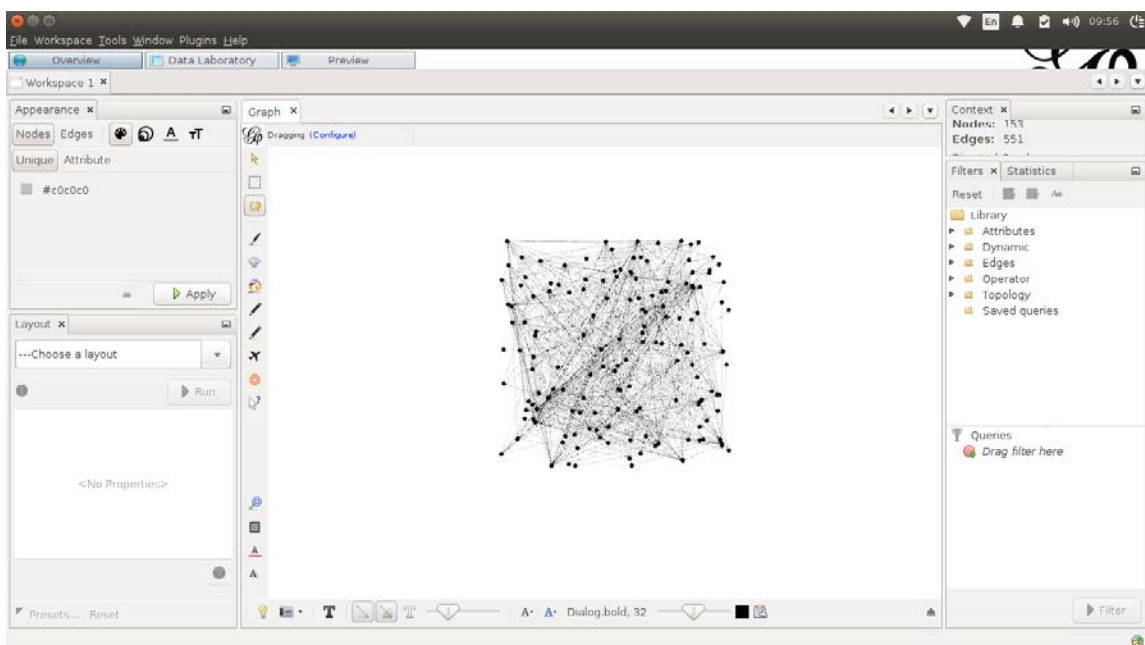
## 1.1. Eksplorasi Gephi

Gephi adalah salah satu perangkat lunak yang digunakan untuk melakukan visualisasi terhadap graf. Berikut adalah tampilan dasar dari Gephi.

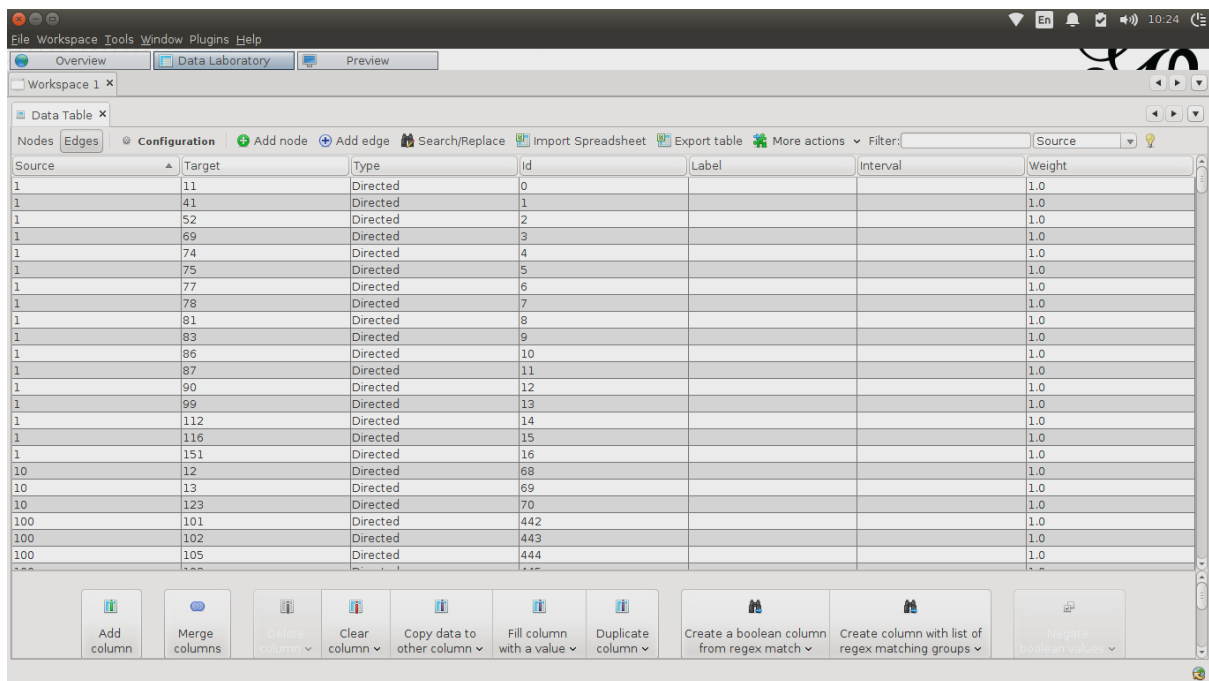
Ada 3 menu utama pada Gephi, yaitu:

1. Overview. Pada menu ini akan dilihat bagaimana *graph* tersebut divisualisasikan secara keseluruhan. Pada bagian ini, kita dapat menentukan apa yang menjadi label dari node dan edge, besar kecil dari node dan edge, penampilan label, layout, dll. Pada menu ini kita juga dapat melihat statistik yang dapat di-ekstrak dari graph tersebut.
2. Data Laboratory. Pada menu ini kita dapat melakukan perubahan terhadap data yang sudah kita *load*. Kita pun dapat melakukan penambahan data graf secara manual. Pada menu ini kita juga dapat melakukan *export* data graf yang saat ini dipakai.
3. Preview. Pada menu ini kita dapat melakukan visualisasi lebih detil terhadap graf.

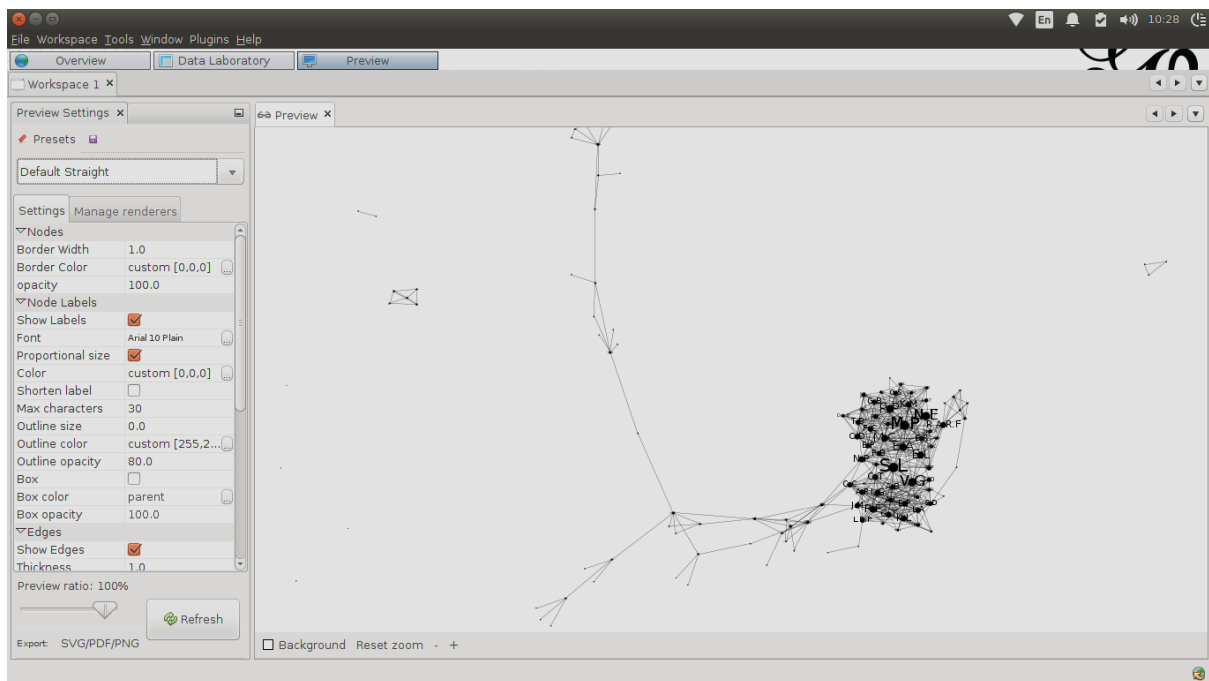
Ketiga tampilan tersebut dapat dilihat pada gambar-gambar berikut.



Gambar 1 Menu Overview pada Gephi

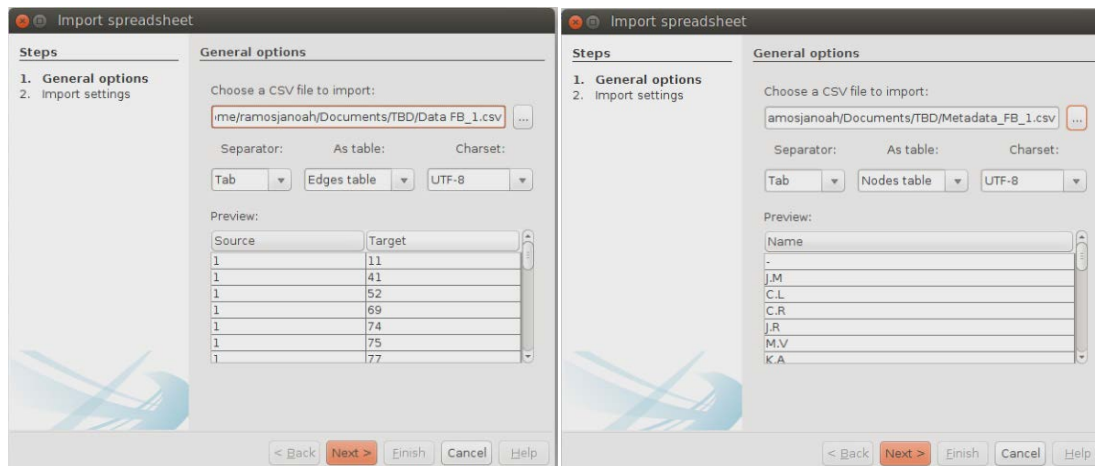


**Gambar 2 Menu Data Laboratory pada Gephi**



**Gambar 3 Menu Preview pada Gephi**

Gephi dapat digunakan untuk mengimport data graf dalam bentuk *edge list* dan *node list*.

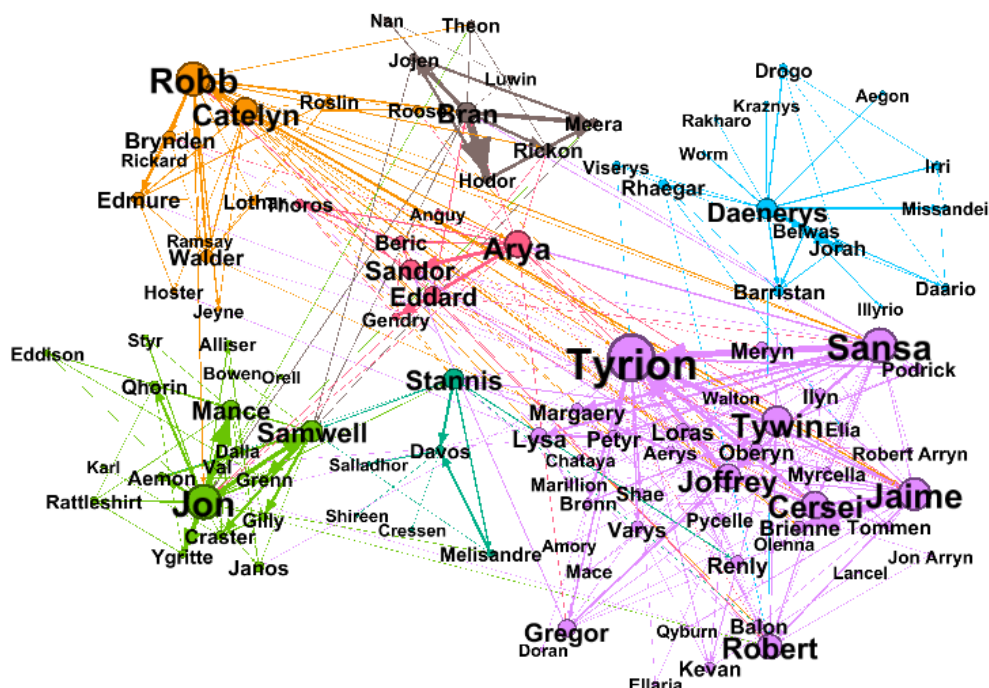


Gambar 4 Import *edge table* dan *node table* pada Gephi

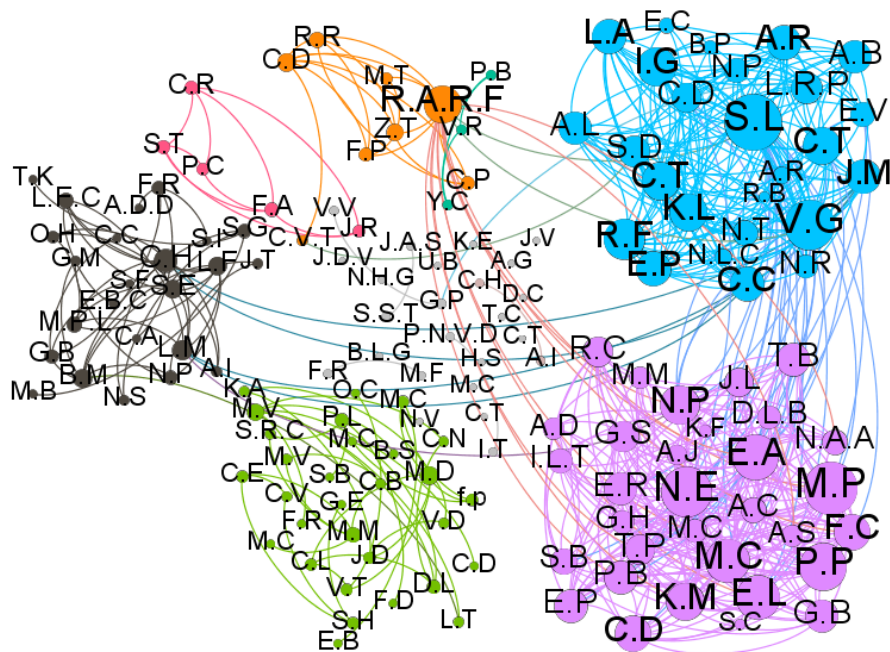
Kemudian data yang sudah di-load dapat di-edit kembali di Menu Data Laboratory.

## 1.2. Pembuatan Visualisasi Menggunakan Gephi

Berikut adalah salah satu contoh hasil visualisasi data graf *Game of Thrones* yang dihasilkan menggunakan gephi.



Gambar 5 Hasil visualisasi menggunakan dataset Network of Thrones



**Gambar 6 Hasil visualisasi pada data Facebook**

Proses-proses yang dilakukan untuk menghasilkan gambar tersebut adalah sebagai berikut:

- Pertama-tama, akan dilakukan *modularity clustering*. Hal ini dilakukan dengan menjalankan *command* 'Modularity' pada tab statistic pada menu overview. Hal ini akan menghasilkan nomor *cluster* dan menjadi atribut dari node
- Kemudian, dilakukan pemisahan node-node berdasarkan cluster. Pemisahan hal ini dilakukan dengan cara menggunakan filter pada menu overview, kemudian pindahkan dengan menggunakan filter attributes → Partition → Modularity Class
- Tahap berikutnya adalah melakukan perubahan terhadap properti dari node dan edge, seperti:
  - o Mengganti warna dari node agar sesuai dengan warna yang ditentukan untuk modularity cluster tersebut
  - o Mengganti ukuran dari cluster dan ukuran dari label (text) dari node agar sesuai dengan 'degree' dari node tersebut
  - o Mengganti ukuran dari edge agar sesuai dengan *weight* dari edge (berlaku untuk data *Game of Thrones*).
- Setelah itu, lakukan layout 'Label Adjust' agar penyusunan Label menjadi lebih rapi.
- Kemudian, lakukan preview dan sesuaikan properti setiap objek agar visualisasi menjadi lebih baik.

## 2. Eksplorasi Frequent Pattern Mining dan Association Rule Mining

Frequent Pattern adalah sebuah pola yang sering terjadi pada fenomena tertentu. Misalnya, pada Market Basket Analysis, frequent pattern yang bisa diekstrak adalah munculnya barang-barang tertentu secara bersamaan pada satu data transaksi.

Pada eksplorasi kali ini, diberikan 3 dataset. Salah satunya adalah transaksi pembelian dengan nama dataset 'Online Retail'. Dilakukan ekstraksi frequent pattern dengan script menggunakan *library Orange3-Association*, yaitu sebuah *library* khusus untuk mencari *frequent pattern* dan *association rule*.

```
3 import orangecontrib.associate.fpgrowth as or3
4 import pandas as pd
5
6 def emptyList():
7     return []
8
9 df_OnlineRetail = pd.read_excel('Online Retail.xlsx')
10
```

Source code 1 import library, inisiasi fungsi dan load data

```
11 # Make ItemList
12
13 dictItemToKey = {}
14 dictKeyToItem = {}
15 dictCounterItem = {}
16 dictItemToDescription = {}
17 firstIndex = df_OnlineRetail.iloc[1][0]
18 IdCounter = 0
19 ItemList = []
20
21 Basket = emptyList()
22 counter = 0
23
24 for index, row in df_OnlineRetail.iterrows():
25     if dictItemToKey.get(row[1], None) == None:
26         dictItemToKey[row[1]] = IdCounter
27         dictItemToDescription[row[1]] = row[2]
28         dictCounterItem[row[1]] = 0
29         dictKeyToItem[IdCounter] = row[1]
30         IdCounter += 1
31     dictCounterItem[row[1]] += 1
32     Basket.append(dictItemToKey[row[1]])
33     if row[0] != firstIndex:
34         ItemList.append(Basket)
35         Basket = emptyList()
36         firstIndex = row[0]
37     counter += 1
38     if counter % 50000 == 0:
39         print(str(counter) + "..")
40
```

Source code 2 Pembuatan *dictionary* untuk dipanggil pada fungsi *library*

```

43 # search for frequent itemset
44 frequentItemSet = list(or3.frequent_itemsets(ItemList, 0.02))
45
46 # search for association Rule
47 associationRuleItemList = or3.association_rules(dict(or3.frequent_itemsets(ItemList, 0.02)),
48                                                    0.001)
49 rules = list(associationRuleItemList)
50

```

**Source code 3 Pemanggilan fungsi fp growth untuk pencarian Frequent Item set dan Association Rule**

Kemudian, dibuat iterasi untuk mendapatkan frequent pattern dengan grid 0,01 dimulai dari 0,01 hingga 0,09. Jumlah pattern yang dihasilkan adalah sebagai berikut:

```

{0.01: 1049,
 0.02: 211,
 0.03: 82,
 0.04: 37,
 0.05: 10,
 0.06: 5,
 0.07: 3,
 0.08: 3,
 0.09: 0}

```

Selain itu, dilihat juga deskripsi barang dari 15 pattern paling *frequent* yang mengandung lebih besar daripada satu *item*, untuk melihat barang apa yang paling sering bersamaan dengan barang lain. Berikut adalah 15 pattern yang paling frequent.

#### FREQUENT PATTERN WITH MORE THAN 1 ITEM

```

[1]
JUMBO BAG PINK POLKADOT
JUMBO BAG RED RETROSPOT
Minimal Support = 796
[2]
GREEN REGENCY TEACUP AND SAUCER
ROSES REGENCY TEACUP AND SAUCER
Minimal Support = 727
[3]
JUMBO BAG RED RETROSPOT
JUMBO STORAGE BAG SUKI
Minimal Support = 701
[4]
JUMBO SHOPPER VINTAGE RED PAISLEY
JUMBO BAG RED RETROSPOT
Minimal Support = 660
[5]
LUNCH BAG RED RETROSPOT
LUNCH BAG SUKI DESIGN
Minimal Support = 633
[6]
LUNCH BAG RED RETROSPOT
LUNCH BAG BLACK SKULL.
Minimal Support = 620
[7]
ALARM CLOCK BAKELIKE RED
ALARM CLOCK BAKELIKE GREEN
Minimal Support = 607
...

```

```

...
[8]
GREEN REGENCY TEACUP AND SAUCER
PINK REGENCY TEACUP AND SAUCER
Minimal Support = 606
[9]
LUNCH BAG PINK POLKADOT
LUNCH BAG RED RETROSPOT
Minimal Support = 590
[10]
JUMBO BAG RED RETROSPOT
JUMBO BAG BAROQUE BLACK WHITE
Minimal Support = 572
[11]
LUNCH BAG SUKI DESIGN
LUNCH BAG BLACK SKULL.
Minimal Support = 569
[12]
LUNCH BAG RED RETROSPOT
JUMBO BAG RED RETROSPOT
Minimal Support = 568
[13]
JUMBO BAG RED RETROSPOT
mail out
Minimal Support = 564
[14]
PINK REGENCY TEACUP AND SAUCER
ROSES REGENCY TEACUP AND SAUCER
Minimal Support = 564
[15]
LUNCH BAG SPACEBOY DESIGN
LUNCH BAG RED RETROSPOT
Minimal Support = 538

```

Dan berikut adalah 10 association rule yang memiliki support terbaik.

```

10 ASSOCIATION RULE WITH GREATER SUPPORT
[1]
JUMBO BAG PINK POLKADOT => JUMBO BAG RED RETROSPOT
Minimal Support = 796
Confidence = 0.6482084690553745
[2]
JUMBO BAG RED RETROSPOT => JUMBO BAG PINK POLKADOT
Minimal Support = 796
Confidence = 0.3738844527947393
[3]
GREEN REGENCY TEACUP AND SAUCER => ROSES REGENCY TEACUP AND SAUCER
Minimal Support = 727
Confidence = 0.6884469696969697
[4]
ROSES REGENCY TEACUP AND SAUCER => GREEN REGENCY TEACUP AND SAUCER
Minimal Support = 727
Confidence = 0.6508504923903312
.....

```



```
.....  
[5]  
JUMBO STORAGE BAG SUKI => JUMBO BAG RED RETROSPOT  
Minimal Support = 701  
Confidence      = 0.5861204013377926  
[6]  
JUMBO BAG RED RETROSPOT => JUMBO STORAGE BAG SUKI  
Minimal Support = 701  
Confidence      = 0.32926256458431186  
[7]  
JUMBO SHOPPER VINTAGE RED PAISLEY => JUMBO BAG RED RETROSPOT  
Minimal Support = 660  
Confidence      = 0.5560235888795282  
[8]  
JUMBO BAG RED RETROSPOT => JUMBO SHOPPER VINTAGE RED PAISLEY  
Minimal Support = 660  
Confidence      = 0.31000469704086425  
[9]  
LUNCH BAG SUKI DESIGN => LUNCH BAG RED RETROSPOT  
Minimal Support = 633  
Confidence      = 0.48580199539524177  
[10]  
LUNCH BAG RED RETROSPOT => LUNCH BAG SUKI DESIGN  
Minimal Support = 633  
Confidence      = 0.394392523364486
```