

# **Analisis Rute Penerbangan Domestik Maskapai Penerbangan Garuda Indonesia**

**Kamila Kaffah, Kezia Sulami**

**Fakultas Ilmu Komputer, Universitas Indonesia**

[kamila.kaffah@ui.ac.id](mailto:kamila.kaffah@ui.ac.id), [kezia.sulami@ui.ac.id](mailto:kezia.sulami@ui.ac.id)

## **A. Latar Belakang Masalah**

PT Garuda Indonesia (Persero) Tbk merupakan salah satu perusahaan milik negara (BUMN) di sektor industri transportasi udara. Dengan jumlah 210 armada, Garuda Indonesia menjangkau 48 rute domestik dan 22 rute internasional [2]. Pada tahun 2013 - 2018, Garuda Indonesia terpilih sebagai Top 10 Global Airlines (10 Besar Maskapai Top Dunia) versi Skytrax. Pada periode yang bersamaan yakni 2014 – 2018 Garuda Indonesia menerima penghargaan ‘The World’s Best Cabin Crew (Awak Kabin Terbaik di Dunia)’ dari Skytrax selama lima tahun berturut-turut [2].

Sebagai maskapai penerbangan terbaik di Indonesia, hingga saat ini belum ada publikasi *paper*/jurnal yang mengangkat topik khusus mengenai keseluruhan rute penerbangan Garuda Indonesia (kode penerbangan: GA). Oleh karena itu, pada artikel ini kami akan membahas topik rute domestik maskapai penerbangan Garuda Indonesia. Artikel ini diharapkan mampu untuk menjawab pertanyaan-pertanyaan terkait rute penerbangan yang bisa dimanfaatkan untuk penelitian di masa mendatang. Pertanyaan yang tercakup pada artikel ini adalah pertanyaan *descriptive analytics*, *inferential analytics*, dan *predictive analytics*. Penjelasan lebih lanjut mengenai ketiga pertanyaan tersebut akan dibahas pada bagian berikutnya.

## B. Pertanyaan Penelitian

Dari latar belakang di atas, dirumuskan pertanyaan penelitian sebagai berikut:

1. (*Descriptive Analytics*) Bandara apa yang paling sibuk pada penerbangan domestik GA?
2. (*Inferential Analytics*) Apakah jalur penerbangan domestik GA cenderung hanya beroperasi di wilayah yang sama? Konteks wilayah di sini adalah 3 wilayah waktu Indonesia, yaitu WIB, WITA, dan WIT.
3. (*Predictive Analytics*) Jalur penerbangan baru apa yang paling mungkin terjadi di bandara Ngurah Rai International Airport (DPS)?

Berikut ini adalah alasan mengapa pertanyaan di atas menarik:

1. Untuk *descriptive analytics*, kami ingin mengetahui bandara mana yang paling sibuk menangani berbagai jalur penerbangan domestik maskapai GA, yakni bandara yang banyak menghubungkan bandara-bandara lainnya melalui jalur penerbangan. Hal ini karena bandara tersebut akan berperan sebagai *transit hub* dimana akan terdapat banyak orang yang berkunjung ke bandara tersebut karena penerbangan mereka, baik transit atau bukan, melalui bandara tersebut.
2. Untuk *inferential analytics*, kami sangat tertarik apakah penerbangan GA lebih cenderung antar WIB saja, antar WITA saja, dan antar WIT saja, atau malah tidak demikian melainkan banyak juga yang dapat dikatakan sebagai “penerbangan jarak jauh” yang bandara asal dan tujuannya berbeda wilayah.
3. Untuk *predictive analytics*, seperti yang kita ketahui, Bali merupakan salah satu pusat pariwisata paling populer di Indonesia. Oleh karena itu kami ingin mengetahui jika akan dibuat jalur penerbangan antara bandara Ngurah Rai International Airport (DPS) di Bali dengan bandara lain, bandara manakah yang merupakan kandidat terbaik? Pembuatan jalur penerbangan yang tepat mungkin saja dapat meningkatkan pemasukan GA sekaligus meningkatkan sektor pariwisata Indonesia karena bisa semakin banyak yang berlibur ke Bali.

### C. Data

Sumber data yang kami gunakan adalah situs resmi maskapai Garuda Indonesia ([www.garuda-indonesia.com](http://www.garuda-indonesia.com)). Pada situs tersebut terdapat data keberangkatan maupun kedatangan penerbangan rute domestik maupun rute internasional Garuda Indonesia secara *real time* sesuai dengan tanggal diaksesnya situs tersebut. Data yang digunakan untuk analisis artikel ini adalah hanya data rute penerbangan domestik pada tanggal 20 April 2021. Data *file* hasil *crawling web* yang telah kami lakukan dapat dilihat pada link berikut:

[https://docs.google.com/spreadsheets/d/1a7rx32YzwYPxHSRDX1K--Zel91UYv6RiUdz\\_OyCFQk8/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1a7rx32YzwYPxHSRDX1K--Zel91UYv6RiUdz_OyCFQk8/edit?usp=sharing).

Keterangan tipe data untuk *dataset* tersebut adalah sebagai berikut:

Kolom	Tipe Data
Kode Bandara Asal	String
Kota Asal	String
Bandara Asal	String
Kode Bandara Tujuan	String
Kota Tujuan	String
Bandara Tujuan	String
Count	Integer
Date	Datetime

Data kolom dari *link file* di atas terdiri dari kode bandara asal, kota asal, bandara asal, kode bandara tujuan, kota tujuan, bandara tujuan, *count*, dan *date*. Lalu, terdapat 70 rute penerbangan. Setiap penerbangan dengan rute tertentu dapat terjadi lebih dari satu kali dan diwakilkan oleh atribut *count*. Total jumlah penerbangan adalah 108 penerbangan. Kami hanya mengambil data pada 20 April 2021 karena jumlah data yang diambil sudah cukup walaupun hanya mengambil data satu hari. Selain itu, rute dan jumlah penerbangan pada hari lainnya kurang lebih sama seperti data pada 20 April 2021.

Data-data rute penerbangan tersebut dapat direpresentasikan dalam bentuk graf. Sekilas terlihat bahwa data atribut “kode bandara asal” dan data atribut “kode bandara tujuan” merupakan *node* pada graf. Lalu, *edge* pada graf adalah rute penerbangan dari *node* “kode bandara asal” menuju *node* “kode bandara tujuan” sehingga membentuk graf berarah. Namun, setelah kami analisis lebih jauh lagi, graf yang terbentuk dari data ini bisa direpresentasikan sebagai graf tidak berarah dan tidak berbobot.

Hasil *crawling* data pada link di atas adalah data *departure* rute domestik, yaitu data keberangkatan dari “bandara asal” menuju “bandara tujuan”. Ketika *crawling* data, kami tidak memasukkan data *arrival* rute domestik karena data tersebut merupakan data yang sama dengan data *departure*. Kami berpikir hal itu akan menyebabkan *redundant* pada data yang menyebabkan penambahan *count* sejumlah 2 kali lipat untuk setiap data rute penerbangan yang didapatkan.

Jika kita merepresentasikan data dengan sudut pandang yang berbeda, kita bisa melihat satu data jalur penerbangan yang didapatkan bisa memuat dua informasi. Kedua informasi tersebut adalah bandara pertama dan bandara kedua (sebelumnya bandara asal dan bandara tujuan). Dengan informasi tersebut, kita dapat merepresentasikan *edge* graf dengan definisi “terdapat rute penerbangan domestik antara bandara pertama dengan bandara kedua”. Berdasarkan hal tersebut, kami bisa melihat graf sebagai graf tidak berarah. Dari 70 *edge* berarah yang ada, setelah kami abaikan arahnya maka menjadi 44 *edge* tidak berarah karena, sebagai contoh, *edge* dari *node* A ke *node* B dan *edge* dari *node* B ke *node* A dianggap sama.

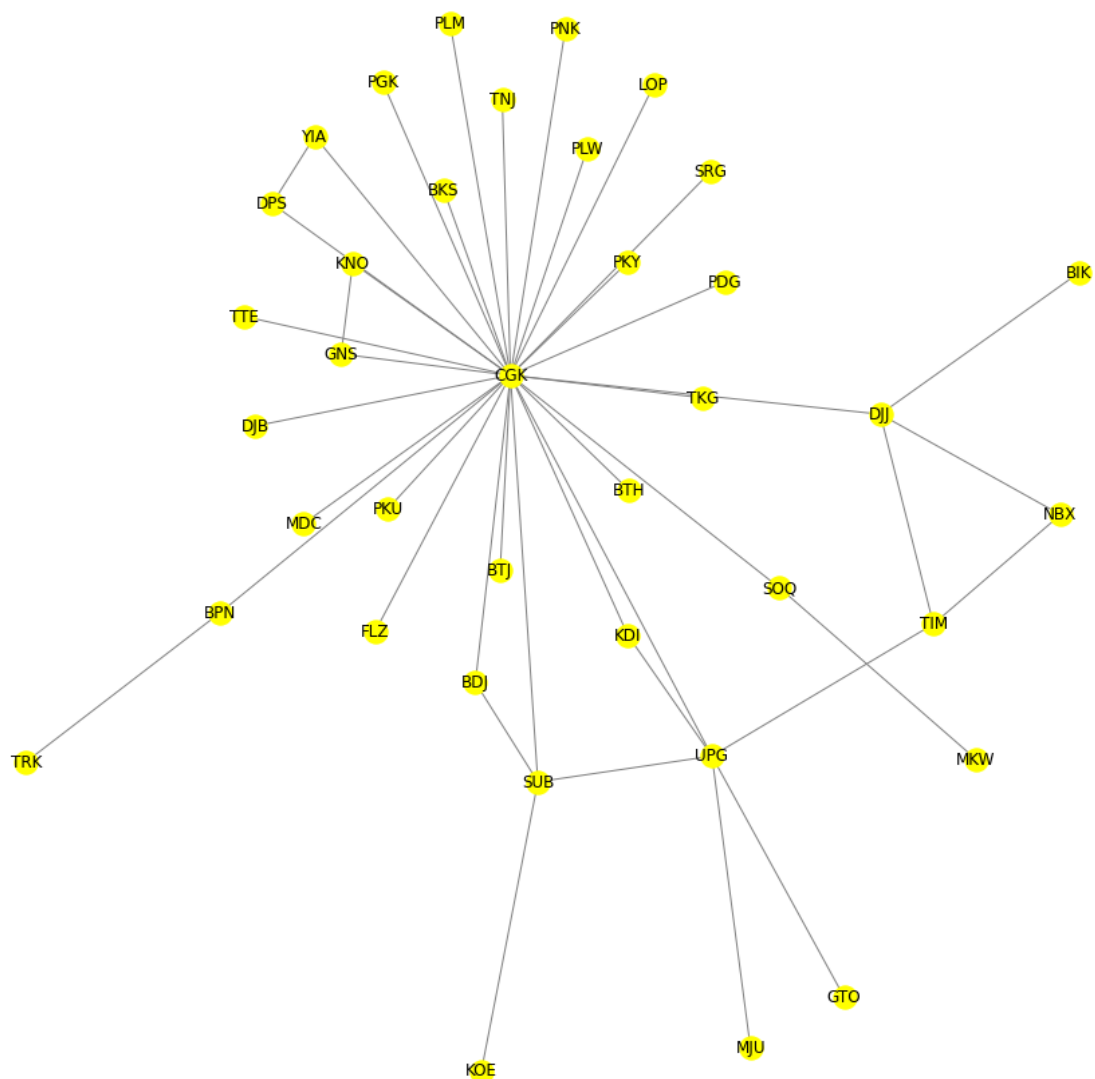
Lalu, definisi bobot pada graf ini akan lebih tepat jika bobot tersebut merupakan jarak atau panjang lintasan penerbangan antara bandara satu dengan bandara kedua. Saat ini, kami tidak memiliki data jarak antar bandara sehingga graf yang terbentuk merupakan graf tidak berbobot.

Keterangan tipe data untuk *dataset* yang telah diolah adalah sebagai berikut:

Kolom	Tipe Data
Kode Bandara 1	String
Kode Bandara 2	String

Ini karena kami hanya memerlukan kode bandaranya saja untuk menggambar struktur graf dan melakukan analisis berdasarkan pertanyaan penelitian yang telah kami susun.

Kesimpulannya adalah data kolom “bandara asal” digunakan sebagai *node* “bandara pertama” dan data kolom “bandara tujuan” digunakan sebagai *node* “bandara kedua”. *Edge* antara *node* “bandara pertama” dan *node* “bandara kedua” merepresentasikan “terdapat rute penerbangan domestik antara bandara pertama dan bandara kedua”. Selanjutnya, graf yang terbentuk dari data adalah graf tidak berarah dan tidak berbobot. Berikut adalah gambar graf yang dihasilkan dari data yang kami miliki.



## D. Metodologi

### 1. Melakukan studi literatur dari *slide* perkuliahan

Berdasarkan pertanyaan penelitian yang telah kami rumuskan, kami perlu meninjau kembali *slide* perkuliahan materi *betweenness centrality* (untuk *descriptive analytics*), *homophily* (untuk *inferential analytics*), dan *link prediction* (untuk *predictive analytics*). Materi-materi tersebut merupakan materi yang paling terkait dengan pertanyaan-pertanyaan penelitian kami dan dapat membantu kami menjawab pertanyaan-pertanyaan penelitian tersebut. Materi-materi tersebut berada pada *slide* nomor 4, 6, dan 8 berturut-turut.

### 2. Melakukan studi literatur dari dokumentasi *library* networkx (dengan Python)

- Kami mempelajari *method* networkx untuk menghitung *betweenness centrality* yaitu `betweenness_centrality()`, tercantum di referensi [3]
- Kami melakukan pencarian dengan kata kunci “networkx homophily” dan menemukan *method* `attribute_assortativity_coefficient()` dan `attribute_mixing_dict()` yang tercantum di referensi [4] dan [5] berturut-turut. Alasan mengapa *method* berdasarkan atribut ini cocok digunakan adalah karena kami menghitung kecenderungan struktur graf yang dipengaruhi oleh suatu atribut, yaitu *timezone* setiap bandara.

Kami mempelajari hal baru yaitu mengenai *assortativity* yang serupa dengan *homophily*. *Attribute assortativity* mengembalikan konstanta yang menyatakan struktur graf cenderung dipengaruhi oleh suatu atribut (di sini *timezone*) atau tidak. Konstanta tersebut bernilai antara -1 s/d 1 dan maknanya serupa dengan korelasi. 1 artinya berkorelasi positif, -1 artinya berkorelasi negatif, dan 0 artinya tidak berkorelasi.

- Kami menggunakan *link prediction* dengan *preferential attachment* menggunakan *method* `preferential_attachment()` sesuai referensi [6]

### 3. Berdiskusi dengan tim pengajar mata kuliah Analitika Media Sosial

Kami berdiskusi dengan dosen Bapak Rahmad Mahendra melalui forum SCELE dan asisten dosen Ariq Naufal Satria melalui *personal chat* di

Telegram. Diskusi meliputi pertanyaan penelitian mengenai graf penerbangan GA kami.

## E. Metode yang Digunakan

Rute penerbangan domestik yang ada pada maskapai Garuda Indonesia GA dapat membentuk suatu jaringan yang direpresentasikan dengan graf. Pada graf tersebut, *nodes* mewakili bandara GA/kota pada bandara tersebut dan *edges* mewakili jalur penerbangan antar bandara GA/kota pada bandara. Dengan menggunakan graf tersebut, kami dapat menjawab pertanyaan yang diajukan menggunakan berbagai macam metode teknis.

Dalam menentukan metode teknis yang akan digunakan, kami menyesuaikan jenis pertanyaan dengan algoritma yang tepat. Terdapat tiga jenis pertanyaan yang dibahas pada artikel ini, yaitu *descriptive*, *inferential*, dan *predictive analytics*. Berikut adalah penjelasan untuk penentuan metode terhadap masing-masing pertanyaan.

### 1. Metode untuk pertanyaan *descriptive analytics*

Pertanyaan *descriptive analytics* yang diajukan adalah “Bandara mana yang paling sibuk pada penerbangan domestik GA?”. Jika dianalisis secara teknis, pertanyaan tersebut menentukan *centrality* pada suatu graf. *Centrality* mendeskripsikan seberapa penting suatu *node* (pada konteks ini *node* adalah bandara/kota) pada suatu jaringan (pada konteks ini jaringan adalah rute penerbangan domestik GA yang membentuk graf). Banyak cara untuk menentukan *centrality* suatu graf dan pada artikel ini kami menggunakan algoritma *betweenness centrality*.

Algoritma *betweenness centrality* mempertimbangkan seberapa penting/intens suatu *node* dalam keterhubungannya dengan *node* lain. Dengan kata lain, dalam menentukan bandara mana yang paling sibuk pada penerbangan GA bisa dilihat seberapa intensitas suatu bandara (*nodes*) pada jalur penerbangan domestik GA menggunakan *betweenness centrality*. Dalam menghitung nilai *betweenness centrality*, kami menggunakan bantuan *networkx library* dari

Python. Berikut adalah cuplikan kode kami ketika menghitung *betweenness centrality*.

```
import operator

# find betweenness centralities
betweenness centrality = nx.betweenness centrality(G)

# get the items and make a list
betweenness centrality = list(betweenness centrality.items())

# sort by centrality values, descending
betweenness centrality.sort(key=lambda data : data[1], reverse=True)

# show 10 betweenness centrality (the rest has value of 0)
betweenness centrality[:10]
```

Pada cuplikan kode di atas, kode akan menghasilkan 10 nilai *betweenness centrality* yang terbesar. Semakin besar nilai *betweenness centrality* pada suatu node bandara, menandakan semakin *central* node bandara tersebut. Dengan kata lain, *node* bandara yang memiliki nilai *betweenness centrality* terbesar adalah bandara yang paling banyak operasi penerbangannya.

## 2. Metode untuk pertanyaan *inferential analytics*

Pertanyaan *inferential analytics* yang diajukan adalah “Apakah jalur penerbangan domestik GA cenderung hanya beroperasi di wilayah yang sama?” Konteks wilayah di sini adalah 3 wilayah waktu Indonesia, yaitu WIB, WITA, dan WIT. Pertanyaan tersebut berkaitan dengan faktor kontekstual (*timezone*) pada rute penerbangan domestik GA. Dalam menjawab pertanyaan tersebut, kami menjawab dengan menggunakan dua algoritma, yaitu assortativity dan Homophily.

Dengan menggunakan assortativity dan Homophily, kami dapat menentukan apakah rute penerbangan domestik GA terbentuk karena faktor kontekstual *timezone* Indonesia atau tidak. Pada artikel ini, kami menggunakan *networkx library* Python untuk menghitung nilai assortativity dan Homophily dari graf yang terbentuk dari rute penerbangan domestik GA. Berikut adalah cuplikan kode perhitungan tersebut.



## ASSORTATIVITY

- a. Dari kode bandara yang ada pada data, kami mencari terlebih dahulu *timezone* berdasarkan wilayah provinsi bandara tersebut.

```
# add attribute timezone to nodes

timezone = {
    'KDI': 'WITA', 'TTE': 'WIT', 'KNO': 'WIB', 'BDJ': 'WITA', 'BIK': 'WIT',
    'TNJ': 'WIB', 'BPN': 'WITA', 'MJU': 'WITA', 'BTJ': 'WIB', 'PNK': 'WIB',
    'DJB': 'WIB', 'PKU': 'WIB', 'SOQ': 'WIT', 'KOE': 'WITA', 'TRK': 'WITA',
    'PLW': 'WITA', 'BTH': 'WIB', 'PKY': 'WIB', 'DJJ': 'WIT', 'LOP': 'WITA',
    'PGK': 'WIB', 'UPG': 'WIB', 'SUB': 'WIB', 'SRG': 'WIB', 'DPS': 'WITA',
    'GNS': 'WIB', 'FLZ': 'WIB', 'BKS': 'WIB', 'TIM': 'WIT', 'NBX': 'WIT',
    'PLM': 'WIB', 'MKW': 'WIT', 'PDG': 'WIB', 'TKG': 'WIB', 'MDC': 'WITA',
    'CGK': 'WIB', 'GTO': 'WITA', 'YIA': 'WIB'
}

nx.set_node_attributes(G, timezone, 'timezone')
```

- b. Perhitungan nilai *assortativity* dapat dilakukan dengan menggunakan fungsi `attribute_assortativity_coefficient(Graph, 'factor')` dari *networkx library*

```
nx.attribute_assortativity_coefficient(G, 'timezone')
```

*Attribute assortativity* mengembalikan nilai konstanta  $r$  yang menyatakan struktur graf cenderung dipengaruhi oleh suatu atribut (di sini *timezone*) atau tidak. Konstanta tersebut didasari oleh *Pearson Correlation* yang bernilai dari -1 sampai 1 [1]. Jika konstanta  $r$  bernilai -1 maka dapat dikatakan *perfect disassortative* atau bandara terhubung namun tidak berdasarkan *timezone*. Lalu, jika konstanta  $r$  bernilai 0 maka dapat dikatakan setiap bandara terhubung secara random. Terakhir, jika konstanta bernilai 1 dapat dikatakan *perfect assortative* atau bandara terhubung satu sama lain berdasarkan *timezone*.

## HOMOPHILY

- a. Kode untuk mencari nilai *fraction* pada Homophily.

```
from collections import Counter
Counter(timezone.values())
```

```
wib = 20/38
wit = 7/38
wita = 11/38

# the expected fraction of cross-timezone edges
2*wib*wit + 2*wib*wita + 2*wit*wita
```

b. Kode untuk menghitung nilai aktual pada graf

```
In [14]: nx.attribute_mixing_dict(G, 'timezone')
Out[14]: {'WIB': {'WIB': 42, 'WIT': 4, 'WITA': 13},
          'WIT': {'WIB': 4, 'WIT': 10},
          'WITA': {'WIB': 13, 'WITA': 2}}

In [15]: # the actual cross-timezone edges. 4 is WIB-WIT and 13 is WIB-WITA
(4 + 13) / len(df)
```

Pada algoritma ini, jika nilai aktual lebih kecil daripada nilai *fraction* maka dapat dikatakan graf bersifat Homophily.

### 3. Metode untuk pertanyaan *predictive analytics*

Pertanyaan *predictive analytics* yang diajukan adalah “Jalur penerbangan baru apa yang paling mungkin terjadi di bandara Ngurah Rai International Airport (DPS)?”. Pertanyaan tersebut dapat dijawab dengan menggunakan konsep *link prediction*. Dalam menentukan *link prediction* terdapat dua metode yang bisa dipakai, yaitu *node neighborhood-based method* dan *methods based on paths between nodes*.

Metode yang kami pilih adalah *node neighborhood-based method*. Hal itu dikarenakan data yang kami miliki tidak memiliki atribut jarak antara bandara satu dengan bandara lainnya sehingga kami tidak bisa menggunakan *methods based on paths between nodes*. Berdasarkan materi *link prediction* pada slide perkuliahan, *methods based on paths between nodes* membutuhkan nilai *shortest paths*. Pada konteks artikel ini, *shortest path* tidak bisa didapatkan hanya dari jumlah rute (*edge*) antara bandara (*node*) satu dengan bandara (*node*) lainnya tetapi butuh data *real* jarak antar bandara.

Menentukan *link prediction* dengan *node neighborhood-based method* dapat dilakukan dengan berbagai macam algoritma. Kali ini, kami memilih algoritma *Preferential Attachment*. Perhitungan algoritma tersebut dapat

dilakukan dengan menggunakan Networkx *library* Python. Berikut adalah cuplikan kode perhitungan tersebut.

```
# find link predictions
link_prediction = list(nx.preferential_attachment(G))

# sort by preferential attachment value, descending
link_prediction.sort(key=lambda data : data[2], reverse=True)

cnt = 0

# show 5 best link predictions containing airport DPS
for route in link_prediction:
    if route[0] == 'DPS' or route[1] == 'DPS':
        print(route)

        cnt += 1
        if cnt == 5:
            break
```

Cuplikan kode di atas menggunakan fungsi `preferential_attachment(Graph)` yang akan mengembalikan nilai *tuple* (bandara 1, bandara 2, *value*). Semakin besar *value* pada *tuple* tersebut, semakin besar pula kemungkinan akan terjadi lintasan baru antara bandara 1 dan bandara 2.

## F. Hasil dan Analisis

### 1. Descriptive Analytics

Dalam menemukan jawaban untuk pertanyaan *descriptive analytics* “Bandara mana yang paling sibuk pada penerbangan domestik GA?” kami menggunakan algoritma *betweenness centrality*. Kami mendapatkan nilai *betweenness centrality* dengan bantuan *networkx library* Python. Berikut adalah cuplikan kode beserta hasilnya setelah kode dijalankan.

```
In [10]: import operator

# find betweenness centralities
betweenness centrality = nx.betweenness centrality(G)

# get the items and make a list
betweenness centrality = list(betweenness centrality.items())

# sort by centrality values, descending
betweenness centrality.sort(key=lambda data : data[1], reverse=True)

# show 10 biggest betweenness centrality (the rest has value of 0)
betweenness centrality[:10]
```

```
Out[10]: [('CGK', 0.9379379379379379),
          ('UPG', 0.13613613613613612),
          ('DJJ', 0.11911911911911913),
          ('SUB', 0.056806806806806796),
          ('BPN', 0.05405405405405406),
          ('SOQ', 0.05405405405405406),
          ('TIM', 0.01126126126126126),
          ('BDJ', 0.0),
          ('BIK', 0.0),
          ('BKS', 0.0)]
```

*Output* yang dihasilkan dari kode merupakan *list tuple* dengan format (kode bandara, nilai *betweenness centrality*). Semakin besar nilai *betweenness centrality* pada suatu *node* bandara, menandakan semakin *central node* bandara tersebut. Dengan kata lain, *node* bandara yang memiliki nilai *betweenness centrality* terbesar adalah bandara yang paling banyak operasi penerbangannya. Berdasarkan *output* di atas, diketahui bahwa *node* dengan nilai *betweenness centrality* terbesar adalah bandara dengan kode CGK dan nilai *betweenness centrality*-nya adalah 0.9379379379379379. Dengan kata lain, bandara yang paling sibuk pada penerbangan domestik GA adalah bandara **Soekarno Hatta International Airport (CGK)**.

## 2. Inferential Analytics

Apakah jalur penerbangan GA cenderung hanya beroperasi di wilayah yang sama? Konteks wilayah disini adalah 3 wilayah waktu Indonesia, yaitu WIB, WITA, dan WIT.

Nilai *assortativity* berdasarkan atribut wilayah waktu (*timezone*) adalah **0.22123893805309733**

```
[ ] nx.attribute_assortativity_coefficient(G, 'timezone')  
  
0.22123893805309733
```

- Nilai *attribute assortativity coefficient* = sekitar 0.22 artinya struktur graf rute penerbangan domestik GA **cenderung dipengaruhi oleh atribut *timezone***, karena nilainya tidak 0, namun pengaruhnya tidak terlalu kuat karena nilai maksimal jika pengaruhnya kuat adalah 1. Pengaruh di sini maksudnya jalur penerbangan GA cenderung beroperasi di wilayah waktu Indonesia yang sama (WIB, WITA, dan WIT).
- Selanjutnya, kami juga melakukan metode lain yaitu *Homophily Test* untuk menyimpulkan graf penerbangan GA *homophily* terhadap *timezone* atau tidak.

Pertama, kami menghitung *expected fraction of cross-timezone edges*, yaitu ekspektasi proporsi banyaknya *edge* yang menghubungkan 2 *node* dimana 2 *node* tersebut merepresentasikan 2 bandara pada *timezone* yang berbeda.

```
from collections import Counter  
Counter(timezone.values())  
  
Counter({'WIB': 20, 'WIT': 7, 'WITA': 11})  
  
wib = 20/38  
wit = 7/38  
wita = 11/38  
  
# the expected fraction of cross-timezone edges  
2*wib*wit + 2*wib*wita + 2*wit*wita  
  
0.6052631578947368
```

Terdapat 38 bandara unik dengan 20 di antaranya terletak pada wilayah WIB, 7 di antaranya terletak pada WITA, dan 11 di antaranya terletak pada WIT. Oleh karena itu, jika dilihat *fraction* dari seluruh bandara untuk  $wib = 20/38$ ,  $wit = 7/38$ , dan  $wita = 11/38$ .

Ekspektasi probabilitas *cross-timezone edges* adalah

$$2 \times wib \times wit + 2 \times wib \times wita + 2 \times wit \times wita$$

Yang bernilai  $\sim 0.61$  seperti pada tangkapan layar di atas.

Selanjutnya, kami menghitung kenyataannya (*actual*) berapa proporsi *edge* yang *cross-timezone*.

```
nx.attribute_mixing_dict(G, 'timezone')
```

```
{'WIB': {'WIB': 42, 'WIT': 4, 'WITA': 13},
 'WIT': {'WIB': 4, 'WIT': 10},
 'WITA': {'WIB': 13, 'WITA': 2}}
```

```
# the actual cross-timezone edges. 4 is WIB-WIT and 13 is WIB-WITA
(4 + 13) / len(df)
```

```
0.38636363636363635
```

```
len(df)
```

```
44
```

Dari data di atas, diketahui bahwa terdapat 44 *edge* dengan persebaran berikut:

	WIB	WIT	WITA
WIB	21	-	-
WIT	4	5	-
WITA	13	0	1

(Untuk WIB-WIT sama saja dengan WIT-WIB, dst., sehingga diberi tanda -)

Adapun totalnya sudah sesuai yaitu  $21 + 4 + 5 + 13 + 0 + 1 = 44$  *edge*.

*“Homophily Test: If the fraction of cross-gender edges is significantly less than  $2pq$ , then there is evidence for homophily.” [7]*

Karena ekspektasi *fraction* = 0.61 sedangkan nilai *actual* = 0.38, termasuk *significantly less* berdasarkan rumusan di atas. Terbukti **terjadi homophily** berdasarkan atribut *timezone* dan jalur penerbangan domestik GA cenderung hanya beroperasi di *timezone* yang sama yaitu antar WIB, WIT, dan WITA.

### 3. Predictive Analytics

Dalam menemukan jawaban untuk pertanyaan *predictive analytics* “Jalur penerbangan baru apa yang paling mungkin terjadi di bandara Ngurah Rai International Airport (DPS)?” kami menggunakan metode *node neighborhood-based*, dan algoritma *Preferential Attachment*. Kami mendapatkan nilai *Preferential Attachment* dengan bantuan *networkx library* Python. Berikut adalah cuplikan kode beserta hasilnya setelah kode dijalankan.

```
In [17]: # find link predictions
link_prediction = list(nx.preferential_attachment(G))

# sort by preferential attachment value, descending
link_prediction.sort(key=lambda data : data[2], reverse=True)

cnt = 0

# show 5 best link predictions containing airport DPS
for route in link_prediction:
    if route[0] == 'DPS' or route[1] == 'DPS':
        print(route)

        cnt += 1
        if cnt == 5:
            break

('DPS', 'UPG', 12)
('DPS', 'SUB', 8)
('DPS', 'DJJ', 8)
('DPS', 'TIM', 6)
('DPS', 'SOQ', 4)
```

*Output* yang dihasilkan dari kode merupakan *tuple* yang berisi bandara pertama, bandara kedua, dan nilai yang didapatkan. Dari informasi tersebut, semakin besar nilai yang didapatkan, semakin besar pula kemungkinan terjadinya rute penerbangan baru antara bandara pertama dan bandara kedua.

Berdasarkan *output* di atas diketahui bahwa nilai *preferential attachment* yang paling besar adalah 12 yang dimiliki oleh bandara kode ‘DPS’ dan bandara kode ‘UPG’. Dengan kata lain, jalur penerbangan baru yang paling mungkin terjadi di bandara Ngurah Rai International Airport (DPS) adalah bandara **Sultan Hasanuddin Airport (UPG), Ujung Pandang, Makassar**.

## G. Kesimpulan

Pertanyaan *analytics* yang dijawab pada artikel ini ada tiga, yaitu (1) Bandara apa yang paling sibuk pada penerbangan domestik GA?; (2) Apakah jalur penerbangan domestik GA cenderung hanya beroperasi di wilayah yang sama? Konteks wilayah di sini adalah 3 wilayah waktu Indonesia, yaitu WIB, WITA, dan WIT; dan (3) Jalur penerbangan baru apa yang paling mungkin terjadi di bandara Ngurah Rai International Airport (DPS)?.

Dalam menjawab ketiga pertanyaan tersebut, kami menggunakan berbagai metode dan algoritma yang disesuaikan dengan jenis soalnya. Pada pertanyaan pertama, kami menentukan bandara yang paling sibuk dengan *betweenness centrality*. Pada pertanyaan kedua, kami menjawab pertanyaan dengan konsep Homophily. Terakhir, pada pertanyaan ketiga, kami menjawab pertanyaan dengan konsep *link prediction*, metode *node neighborhood-based*, dan algoritma *Preferential Attachment*. Berikut adalah tabel hasil jawaban dari ketiga pertanyaan yang diajukan berdasarkan penjelasan pada bagian Hasil dan Analisis.

Pertanyaan <i>Analytics</i>	Jawaban
Bandara apa yang paling sibuk pada penerbangan domestik GA?	Dengan menggunakan algoritma <i>betweenness centrality</i> , didapatkan bandara yang paling sibuk adalah bandara Soekarno Hatta International Airport (CGK)
Apakah jalur penerbangan domestik GA cenderung hanya beroperasi di wilayah	Dengan menggunakan konsep Homophily, didapatkan nilai ekspektasi



<p>yang sama?</p>	<p><i>fraction</i> = 0.61 sedangkan nilai <i>actual</i> = 0.38, termasuk <i>significantly less</i> berdasarkan rumusan di atas. Terbukti terjadi <i>homophily</i> berdasarkan atribut <i>timezone</i> dan jalur penerbangan domestik GA cenderung hanya beroperasi di <i>timezone</i> yang sama yaitu antar WIB, WIT, dan WITA.</p> <p>Sama juga ketika kami menggunakan algoritma assortativity, didapatkan nilai <i>attribute assortativity coefficient</i> = sekitar 0.22 (<i>positive</i>) menunjukkan adanya Homophily.</p>
<p>Jalur penerbangan baru apa yang paling mungkin untuk terjadi di bandara Ngurah Rai International Airport (DPS)</p>	<p>Dengan menggunakan algoritma <i>Preferential Attachment</i>, jalur penerbangan baru yang paling mungkin terjadi di bandara Ngurah Rai International Airport (DPS) adalah bandara Sultan Hasanuddin Airport (UPG), Ujung Pandang, Makassar</p>

## Daftar Pustaka

- [1] Kasthuriratna, Dharshana et al. (Desember, 2014). *Node Assortativity in Complex Networks: An Alternative Approach*. Diakses 10 Mei, 2021 dari ResearchGate:  
[https://www.researchgate.net/publication/316472185\\_Node\\_Assortativity\\_in\\_Complex\\_Networks\\_An\\_Alternative\\_Approach](https://www.researchgate.net/publication/316472185_Node_Assortativity_in_Complex_Networks_An_Alternative_Approach)
- [2] Yuniarto, Topan. (17 Februari, 2021). PT Garuda Indonesia (Persero) Tbk. Diakses 10 Mei, 2021 dari Kompas Pedia:  
<https://kompaspedia.kompas.id/baca/profil/lembaga/pt-garuda-indonesia-persero-tbk>
- [3] Networkx.algorithms centrality.betweenness centrality. (22 Agustus, 2020). Diakses 10 Mei, 2021 dari Dokumentasi NetworkX:  
<https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms centrality.betweenness centrality.html#networkx.algorithms centrality.betweenness centrality>
- [4] Networkx.algorithms.assortativity.attribute assortativity coefficient. (22 Agustus, 2020). Diakses 10 Mei, 2021 dari Dokumentasi NetworkX:  
<https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.assortativity.attribute assortativity coefficient>
- [5] Networkx.algorithms.assortativity.attribute mixing dict. (22 Agustus, 2020). Diakses 10 Mei, 2021 dari Dokumentasi NetworkX:  
<https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.assortativity.attribute mixing dict>
- [6] Networkx.algorithms.link prediction.preferential attachment. (22 Agustus, 2020). Diakses 10 Mei, 2021 dari Dokumentasi NetworkX:  
<https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.link prediction.preferential attachment>
- [7] Slide perkuliahan Analitika Media Sosial, Semester Genap 2020/2021