

Implementasi *Closeness Centrality* dalam Analisis Penyanyi Indonesia di DBpedia Indonesia

Nur Aini Rakhmawati⁽¹⁾, Ahmad Naufal Muzakki⁽²⁾, Luthfi Lazuardi^{(3)*}

Sistem Informasi, Fakultas Teknologi Informasi dan Komunikasi, Institut Teknologi Sepuluh Nopember, Surabaya

e-mail : nur.aini@is.its.ac.id, {naufal.17052,lazuardiluthfi.17052}@mhs.its.ac.id.

* Penulis korespondensi.

Artikel ini diajukan 22 Mei 2020, direvisi 22 Juli 2020, diterima 24 Juli 2020, dan dipublikasikan 20 Januari 2021.

Abstract

Indonesian singers have a diverse musical genre. They also origin from several regions in Indonesia. This study analyzes the relationships between singers, genre, and origin of the singer using the *Closeness Centrality* algorithm. Data are retrieved by using a SPARQL query on DBpedia Indonesia. The generated graph produces a density value of 0.01. The results of the *Closeness Centrality* calculation shows that more than 50% out of 268 singers had scores above 30%.

Keywords: Indonesian Singers, SPARQL, DBpedia, *Closeness Centrality*, Graph

Abstrak

Penyanyi Indonesia mempunyai aliran musik yang beragam dan juga berasal dari berbagai daerah di Indonesia. Penelitian ini menganalisa hubungan antar penyanyi, genre dan daerah asal penyanyi dengan menggunakan algoritma *Closeness Centrality*. Data yang didapat merupakan hasil SPARQL query dari DBpedia Indonesia. Graf yang terbentuk menghasilkan nilai kepadatan 0,01. Hasil dari perhitungan algoritma *Closeness Centrality* terhadap 268 penyanyi didapatkan bahwa 50% lebih penyanyi Indonesia memiliki nilai di atas 30%.

Kata Kunci: Penyanyi Indonesia, SPARQL, DBpedia, *Closeness Centrality*, Graf

1. PENDAHULUAN

Jejaring sosial merupakan struktur yang dibentuk oleh seperangkat *node* yang terhubung dengan ikatan yang berbeda. Pada perspektif jejaring sosial, *node* dapat berupa banyak jenis aktor sosial, seperti individu, kelompok, perusahaan, organisasi. Sementara itu, ikatan juga dapat mewakili koneksi dan hubungan yang berbeda (Zhang & Luo, 2017). Analisis jejaring sosial merupakan studi tentang jejaring sosial yang berbeda, yaitu jaringan yang terdiri dari manusia dan menggunakan konsep teori grafik dan analisis multivariat untuk menganalisis dan memahaminya (Shaikh & Jiabin, 2006). Salah satu konsep yang digunakan untuk melakukan analisis jejaring sosial ini yaitu *centrality principles*. *Centrality principles* yang diterima secara universal seperti *degree*, *closeness*, *betweenness* dan *eigen vector* dapat digunakan untuk memahami dan menganalisis sebuah jaringan (Gaharwar & Shah, 2018).

Pada penelitian ini dibahas terkait analisis sentralitas pada data penyanyi Indonesia yang berasal dari id.dbpedia.org. Pada analisis yang dilakukan digunakan tiga *node* yaitu *node* nama penyanyi, genre dan juga daerah asal penyanyi. Tujuan dari penelitian ini yaitu untuk menganalisis kedekatan antar penyanyi Indonesia menggunakan *closeness centrality*. *Closeness centrality* merupakan konsep yang sentralitas menjelaskan kedekatan antar *node* dalam jaringan (Kolesnikov et al., 2019). Oleh karena itu *closeness centrality* dapat digunakan untuk mengetahui seberapa besar hubungan antar penyanyi Indonesia berdasarkan asal daerah dan aliran musiknya.

Closeness centrality juga telah banyak digunakan untuk melakukan analisis jejaring sosial pada area atau kasus-kasus lain, seperti menentukan fokus topik pada *linked data* (Abele et al., 2017), melakukan analisa antar entitas berita palsu (Gomes Jr. & Frizzon, 2019), menemukan faktor



yang mempengaruhi peran dan posisi seseorang dalam sebuah tim (Baek & Bae, 2019) dan menganalisis jaringan makanan (Kolesnikov et al., 2019).

1.1. Closeness Centrality

Konsep sentralitas menjelaskan tentang menonjol atau tidaknya suatu *node* dalam jaringan. Sentralitas adalah ukuran dalam grafik yang digunakan dalam analisis jaringan untuk menemukan struktur penting dari *node* dan *edge*. *Closeness centrality* merupakan rata-rata jarak antar *node* dibandingkan dengan *node* lainnya dalam sebuah graf (Abele et al., 2017; Chen et al., 2019; Perwiradewa et al., 2020).

Pada penelitian ini, analisis dilakukan menggunakan jenis pengukuran *closeness centrality*. Rumus di bawah ini merupakan rumus perhitungan *closeness centrality* setiap *node* dalam jaringan (Kolesnikov et al., 2019):

$$C(x, V) = \frac{1}{\sum_{y \in V} d(y, x)} \quad (1)$$

Di mana x merupakan sebuah *node*, V merupakan himpunan vertex, dan $d(y, x)$ merupakan jarak terpendek antara *node* x dan *node* y .

1.2. SPARQL

Sementara semakin banyak data terstruktur yang diterbitkan di web mengikuti prinsip-prinsip *linked data* (Bizer et al., 2011), pertanyaan penting adalah bagaimana seseorang dapat secara efisien mengakses dan meminta informasi yang terus meningkat ini. SPARQL adalah *standard query language* untuk mengambil dan memanipulasi data RDF (Harris et al., 2013). Namun, sebagian besar implementasi SPARQL mengharuskan data tersedia terlebih dahulu, misal, ada di memori utama atau dalam repositori RDF yang dapat diakses melalui titik akhir SPARQL. Meskipun demikian, *linked data* ada di *web* dalam berbagai bentuk, bahkan halaman *web* HTML dapat berisi data RDF (Rakhmawati, 2015), atau data RDF dapat dibuat secara dinamis oleh *web services*.

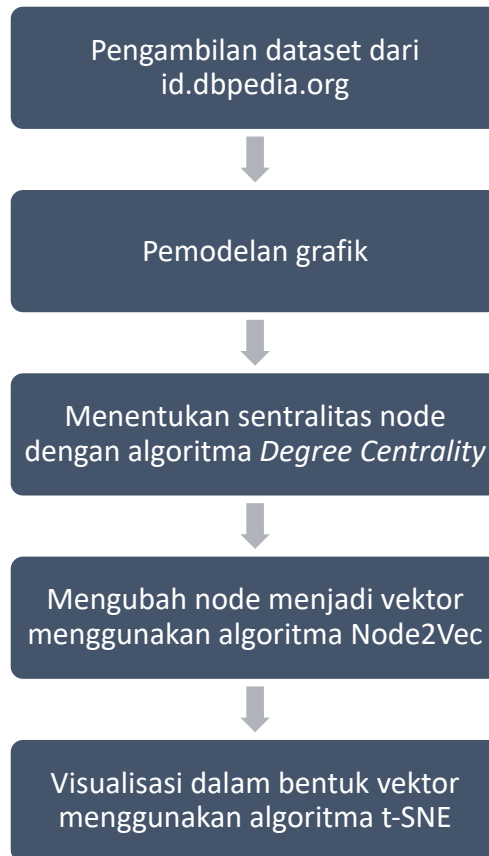
1.3. Node2Vec

Node2Vec (Grover & Leskovec, 2016) merupakan strategi pengambilan sampel lingkungan yang fleksibel yang memungkinkan untuk melakukan interpolasi dengan lancar antara *Breadth First Search* (BFS) dan *Depth First Search* (DFS). Hal ini dilakukan dengan mengembangkan prosedur berjalan acak (*random walks*) yang fleksibel yang dapat menjelajahi lingkungan dalam mode BFS dan DPS (Needham & Hodler, 2019).

2. METODE PENELITIAN

Penelitian dimulai dari pembuatan dataset yang datanya diambil dari id.dbpedia.org. Kemudian dilanjutkan dengan pemodelan grafik yang berisi data-data pada dataset yang diubah menjadi bentuk *node* serta hubungan antar *node-node* tersebut. Kemudian dilanjutkan dengan menentukan sentralitas dari tiap-tiap *node* menggunakan algoritma *degree centrality* dan divisualisasikan kembali dalam bentuk grafik. Data yang sudah dalam bentuk *node* selanjutnya diubah menjadi vektor menggunakan algoritma Node2Vec dan divisualisasikan menggunakan algoritma t-SNE. Alur metodologi penelitian digambarkan pada Gambar 1 berikut.





Gambar 1. Alur metode penelitian.

2.1. Pengambilan Dataset dengan SPARQL

Dataset yang digunakan adalah *dataset* Penyanyi Indonesia berdasarkan DBpedia. Atribut (kolom) yang akan digunakan pada *dataset* ini yaitu nama penyanyi, genre penyanyi, dan asal penyanyi. Data pada kolom-kolom tersebut akan diubah menjadi bentuk *node* yang terhubung dan menjadi sebuah grafik.

Pengambilan *dataset* dilakukan dengan menggunakan *query* SPARQL. Data yang diambil yaitu data penyanyi Indonesia, genre dan asal penyanyi, masing-masing menjadi satu kolom pada *dataset* yang dibuat. Untuk memperoleh data sesuai yang dibutuhkan, didapatkan dengan melakukan *query* seperti berikut.

```

SELECT DISTINCT ?nama ?genre ?asal WHERE {
  ?penyanyi a dbpedia-owl:Person.
  ?penyanyi dbpedia-owl:wikiPageWikiLink
  <http://id.dbpedia.org/resource/Kategori:Penyanyi_Indone
  sia>.
  ?penyanyi rdfs:label ?nama.
  ?penyanyi <http://id.dbpedia.org/property/genre> ?genre.
  ?penyanyi
  <http://id.dbpedia.org/property/birthplace> ?asal.
}
  
```



Dataset hasil *query* di atas kemudian disimpan dalam bentuk atau format CSV di mana nama penyanyi disimpan dalam kolom “nama”, genre disimpan pada kolom “genre” dan asal penyanyi disimpan pada kolom “asal”. Sepuluh data pertama pada *dataset* disajikan pada Tabel 1.

Tabel 1. Hasil CSV dari query SPARQL penyanyi Indonesia.

Nama	Genre	Asal
Acha Septriasa	http://id.dbpedia.org/resource/Pop	Jakarta,
Acha Septriasa	http://id.dbpedia.org/resource/Soundtrack	Jakarta,
Agnes Monica	http://id.dbpedia.org/resource/R&B	Jakarta, Indonesia
Agnes Monica	http://id.dbpedia.org/resource/Pop	Jakarta, Indonesia
Ahmad Albar	http://id.dbpedia.org/resource/Pop	Surabaya, Indonesia
Ahmad Albar	http://id.dbpedia.org/resource/Rock	Surabaya, Indonesia
Andi Meriem Matalatta	http://id.dbpedia.org/resource/Pop	Makassar, Sulawesi Selatan, Indonesia
Anneke GrÃ¶nloh	http://id.dbpedia.org/resource/Big_band	Tondano, Sulawesi Utara, Indonesia
Anneke GrÃ¶nloh	http://id.dbpedia.org/resource/Pop	Tondano, Sulawesi Utara, Indonesia
Agnes Monica	http://id.dbpedia.org/resource/R&B	Jakarta, Indonesia

Data hasil *query* perlu dilakukan pembersihan, sebagaimana contoh di atas Acha Septriasa dan Agnes Monica sama-sama lahir di Jakarta, namun data untuk Agnes Monica terdapat kata Indonesia. Hal ini dilakukan untuk mempermudah pengolahan data. *Full dataset* dan kode dapat diakses pada situs Zenodo (Rakhmawati, 2020).

2.2. Pemodelan Grafik

Dalam penelitian ini, untuk menggambarkan hubungan atribut dalam bentuk grafik, ditetapkan 3 jenis *node* (nama, genre, dan asal) yang masing-masing merupakan atribut pada data penyanyi Indonesia. Setelah melakukan pemodelan grafik kemudian dicari juga kepadatan dari jaringan (*network density*) dari data yang ada.

2.3. Penentuan Closeness Centrality Node

Pada bagian ini dilakukan penentuan sentralitas pada data berdasarkan *degree* untuk setiap *node* yang ada pada grafik menggunakan algoritma *closeness centrality*.

2.4. Menjalankan Node2vec

Dalam penelitian ini, digunakan fungsi *node2vec* Python untuk menggambarkan grafik yang akan dibuat. Parameter *node2vec* yang digunakan pada kasus ini, yaitu:

- Dimension = 20
- Walk_lenght = 16
- Num_walks = 100
- Workers = 2

2.5. Visualisasi dengan Algoritma t-SNE

Untuk memvisualisasikan hasil penelitian, digunakan *tools* Jupyter Notebook dan menggunakan berbagai *library* seperti Matplotlib dan TSNE. Hasil akhirnya merupakan visualisasi grafik *closeness centrality* dan *node2vec*.



Tabel 2. Top 10 node dengan nilai *closeness centrality* tertinggi.

No	Node	Closeness Centrality
1	Pop	54.85%
2	Denada	40.20%
3	Eka Deli	39.71%
4	Erie Suzan	39.62%
5	Andity	38.90%
6	Dessy Fitri	38.90%
7	Elvyn g. masassya	39.82%
8	Rio Silaen	38.67%
9	Chrisye	38.59%
10	vincent ryan rompies	38.59%

3.3. Node2Vec

Graf yang terbentuk ditransformasikan ke dalam bentuk *embedding* agar dapat terlihat kedekatan hubungan antara penyanyi, genre dan daerah asalnya. Berdasarkan parameter yang telah disebutkan sebelumnya, ditampilkan 10 *node* yang mirip dengan *node* genre Dangdut dan didapatkan hasil seperti pada Tabel 3 berikut.

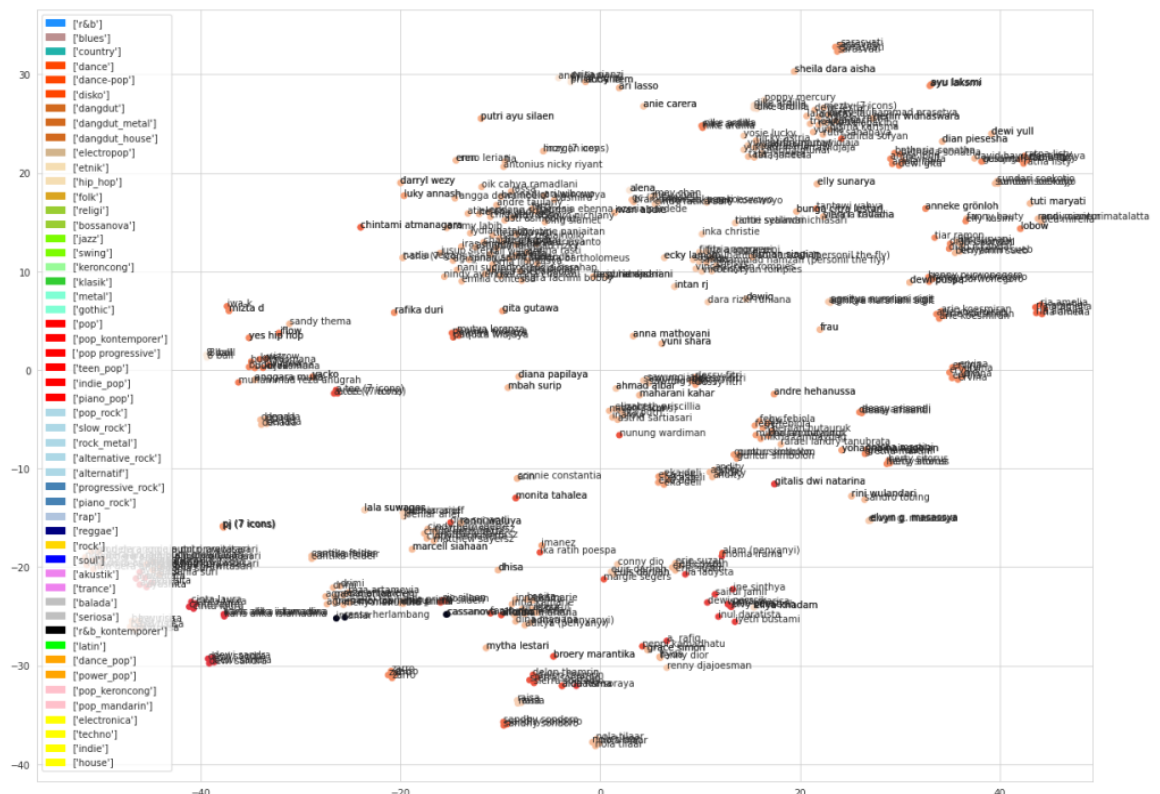
Tabel 3. Sepuluh (10) *node* yang memiliki kemiripan dengan *node* genre Dangdut.

No	Node
1	lia ladysta
2	wawa marisa
3	julia perez
4	saiful jamil
5	elvy sukaesih
6	iyeth bustami
7	Bengkalis
8	nada soraya
9	rhoma irama
10	dewi perssik

3.4. Visualisasi Menggunakan Algoritma t-SNE

Visualisasi hasil *node2vec* menggunakan t-SNE berdasarkan genre penyanyi. Hasil yang ditampilkan pada Gambar 3 di bawah ini merupakan nama-nama penyanyi Indonesia. Pada bagian bawah tengah terdapat penyanyi dangdut seperti Roma Irama, Dewi Persik dan Inul Darastita. Penyanyi *rock* berkumpul di bagian atas tengah. *Node* yang sangat mirip terkait genre dan daerah asalnya akan terlihat bertumpuk, misal penyanyi Rini Wulandari dan Sandro Tobing memiliki aliran musik pop dan berasal dari daerah Medan.





Gambar 3. Visualisasi menggunakan t-SNE.

4. KESIMPULAN

Penelitian ini telah melakukan analisa penyanyi Indonesia yang terdapat pada DBpedia Indonesia dengan menggunakan *closeness centrality*. Lebih dari separuh penyanyi memiliki nilai *closeness centrality* di atas 30%. Dengan penggunaan algoritma *Closeness Centrality* dapat diketahui genre terbanyak yaitu pop. Visualisasi graf *embedding* mampu memperlihatkan hubungan antara penyanyi, aliran dan daerah asalnya.

DAFTAR PUSTAKA

- Abele, A., McCrae, J. P., & Buitelaar, P. (2017). An Evaluation Dataset for Linked Data Profiling. *Lecture Notes in Computer Science*, 1–9. https://doi.org/10.1007/978-3-319-59888-8_1
- Baek, S. I., & Bae, S. H. (2019). The Effect of Social Network Centrality on Knowledge Sharing. *Journal of Service Science Research*, 11(2), 183–202. <https://doi.org/10.1007/s12927-019-0009-2>
- Bizer, C., Heath, T., & Berners-Lee, T. (2011). Linked Data: The Story so Far. In *Semantic Services, Interoperability and Web Applications: Emerging Concepts* (pp. 205–227). IGI Global.
- Chen, J., Jiang, Y., & Yang, M. (2019). Collaborative Recommendation for Scenic Spots based on Degree Centrality. *2019 6th International Conference on Systems and Informatics (ICSAI)*, 611–615. <https://doi.org/10.1109/ICSAI48974.2019.9010437>
- Gaharwar, R. D., & Shah, D. B. (2018). Use of Degree Centrality Principle in Deciding the Future Leader of the Terrorist Network. *International Journal of Scientific Research in Science and Technology (IJSRST)*, 4(9), 303–310.
- Gomes Jr., L., & Frizzon, G. (2019). Fake News and Brazilian politics – temporal investigation based on semantic annotations and graph analysis. *XXXIV Simpósio Brasileiro de Banco de Dados*, 169–174. <https://doi.org/10.5753/sbbd.2019.8818>
- Grover, A., & Leskovec, J. (2016). node2vec: Scalable Feature Learning for Networks.



- Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 855–864. <https://doi.org/10.1145/2939672.2939754>
- Harris, S., Seaborne, A., & Prud'hommeaux, E. (2013). *SPARQL 1.1 Query Language*. W3C.
- Kolesnikov, V., Anikin, V., Mosolova, E., Faizliev, A., Mironov, S., Zemlyanskaya, M., Pleshakov, M., & Sidorov, S. (2019). Food Chain Analysis Based on Graph Centrality Indicators. *Journal of Physics: Conference Series*, 1334, 1–11. <https://doi.org/10.1088/1742-6596/1334/1/012004>
- Needham, M., & Hodler, A. E. (2019). *Graph Algorithms: Practical Examples in Apache Spark and Neo4j* (1st ed.). O'Reilly Media.
- Perwiradewa, A., Rofiif, A. N., & Rakhmawati, N. A. (2020). Visualisasi Pemain Sepak Bola Indonesia pada DBPedia dengan menggunakan Node2Vec dan Closeness Centrality. *Jurnal Buana Informatika*, 11(2), 104–111. <https://doi.org/10.24002/jbi.v11i2.3346>
- Rakhmawati, N. A. (2015). *Semantic Web dan Linked Data* (Yono (ed.); 1st ed.). Sibuku Media.
- Rakhmawati, N. A. (2020). *nurainir/Penyanyi-Centrality: submittojiska*. <https://doi.org/10.5281/ZENODO.3956065>
- Shaikh, M. A., & Jiaxin, W. (2006). Investigative Data Mining: Identifying Key Nodes in Terrorist Networks. *2006 IEEE International Multitopic Conference*, 201–206. <https://doi.org/10.1109/INMIC.2006.358163>
- Zhang, J., & Luo, Y. (2017). Degree Centrality, Betweenness Centrality, and Closeness Centrality in Social Network. *2017 2nd International Conference on Modelling, Simulation and Applied Mathematics (MSAM2017)*, 300–303. <https://doi.org/10.2991/msam-17.2017.68>

