Deteksi Komunitas Spesies Laba-Laba Berdasarkan Negara Menggunakan Algoritma Label Propagation

Imam Mansyur Solichin [1], Bayu Narendra Jati[2], Nur Aini Rakhmawati[3]

[1,2,3,4] Program Studi S1 Sistem Informasi, Institut Teknologi Sepuluh Nopember
[1]imam.18052@mhs.its.ac.id
[2]narendra.18052@mhs.its.ac.id
[3]nur.aini@is.its.ac.id

Abstrak*— Laba-laba merupakan salah satu hewan yang tergolong dalam filum artropoda. Laba-laba merupakan salah satu dari kelompok yang paling krusial bagi ekosistem alami dan pertanian, laba-laba dapat menjadi indikator alami keanekaragaman hewan dalam suatu daerah. Berkaitan dengan hal tersebut, pengelompokkan spesies laba-laba berdasarkan negara habitatnya dapat menjadi hal yang menarik untuk diteliti. Pada penelitian ini algoritma deteksi komunitas digunakan untuk mengelompokkan spesies laba-laba menjadi komunitas yang mendiami negara-negara tertentu, algoritma deteksi komunitas yang digunakan adalah label propagation yang mendeteksi komunitas menggunakan struktur jaringan sebagai panduan, dan tidak memerlukan fungsi tujuan yang telah ditetapkan atau informasi awal tentang komunitas. Pemodelan graf Neo4j digunakan untuk menampilkan hasil algoritma label propagation untuk mendeteksi komunitas spesies laba-laba. Hasil penelitian menunjukkan bahwa algoritma label propagation dapat mengelompokkan komunitas spesies laba-laba sebanyak 159 komunitas. Negara China merupakan negara dengan komunitas spesies laba-laba terbesar di dunia dengan jumlah spesies laba-laba berjumlah 279 spesies, diikuti oleh Jepang dan Amerika Serikat dengan masing-masing 86 spesies laba-laba.

Kata Kunci*— *Latent Dirichlet Allocation*, *Gaming*, analisis topik, topik dominan, konten *youtube

I. PENDAHULUAN

Laba-laba merupakan salah satu hewan yang tergolong dalam filum artropoda. Hewan yang tergolong dalam filum merupakan hewan yang memiliki ruas atau buku pada kakinya, artropoda merupakan filum terbesar dalam kerajaan hewan. Salah satu bagian dari filum ini adalah ordo arachnida, yang merupakan bangsa laba-laba. Laba-laba diperkirakan memiliki 114.275 spesies, dengan jumlah spesies yang sangat beragam tersebut laba-laba dapat ditemukan di hampir seluruh belahan dunia kecuali antartika [1]. Laba-laba merupakan salah satu dari kelompok yang paling krusial bagi ekosistem alami dan pertanian. Laba-laba tersebar di seluruh dunia dengan, meskipun begitu laba-laba sensitif terhadap keadaan lingkungan. Karena itu laba-laba dapat menjadi indikator alami keanekaragaman hewan dalam suatu daerah [2]. Berkaitan dengan hal tersebut, pengelompokkan spesies laba-laba berdasarkan negara habitatnya dapat menjadi hal yang menarik untuk diteliti.

Deteksi komunitas merupakan algoritma yang bekerja dengan cara mengelompokkan node untuk membentuk suatu komunitas [3]. Di sisi lain juga terdapat algoritma *closeness centrality* yang bekerja dengan cara mengukur jarak rata-rata sebuah node ke semua node yang ada. Sebuah node memiliki nilai kedekatan yang tinggi apabila jarak node tersebut mempunyai jarak yang pendek ke semua node yang ada [4]. Penelitian sebelumnya yang menggunakan algoritma *closeness centrality* menganalisis suku yang ada di Indonesia untuk mencari hubungan dari setiap suku [5]. Pada penelitian ini algoritma deteksi komunitas digunakan untuk mengelompokkan spesies laba-laba menjadi komunitas yang mendiami negara-negara tertentu. Algoritma deteksi komunitas dipilih daripada *closeness centrality* karena penelitian ini berfokus untuk melihat pengelompokan spesies dan bukan kedekatan antar spesies. Salah satu jenis algoritma deteksi komunitas adalah *label propagation* yang akan digunakan dalam penelitian ini. Algoritma *label propagation* adalah algoritma cepat untuk menemukan komunitas dalam graf. Algoritma ini mendeteksi komunitas menggunakan struktur jaringan saja sebagai panduannya, dan tidak memerlukan fungsi tujuan yang telah ditentukan sebelumnya atau informasi sebelumnya tentang komunitas [6].

Pada penelitian sebelumnya [7], algoritma deteksi komunitas digunakan untuk mengelompokkan pahlawan yang ada di Indonesia berdasarkan data daerah dan tahun. Terdapat juga penelitian sebelumnya [8] yang menggunakan algoritma deteksi komunitas untuk melakukan analisis jaringan sosial untuk mengetahui pola relasi antar kandidat pemilihan umum.

Penelitian ini dilakukan dengan tujuan agar data persebaran spesies laba-laba di dunia dapat dikelompokkan menjadi komunitas spesies yang mendiami negara-negara tertentu. .

II. TINJAUAN PUSTAKA

A. Neo4j

Neo4j adalah basis data graf open source yang menyimpan data dalam bentuk graf daripada tabel statis. Neo4j tersusun atas 4 dasar yaitu node, relasi, properti dan label. Node adalah elemen data utama, yang terhubung ke node lain dengan relasi. Label digunakan untuk mendeskripsikan node, sementara properti dapat dimiliki oleh node dan relasi yang berupa pasangan key-value [9].

Neo4j mempunyai berbagai keunggulan, diantaranya:

1. ACID transaction pada Neo4j memastikan data sepenuhnya konsisten dan reliabel untuk aplikasi enterprise global.

2. Mendukung konsep banyak pengguna dan kerangka kerja kontrol akses berbasis peran terbaru.
3. Neo4j menyediakan arsitektur plug-in baru untuk membangun dan menerapkan kontrol otentikasi dan otorisasi pengguna sendiri.

Neo4j juga memiliki beberapa kekurangan, diantaranya:

1. Tidak dapat membagikan Neo4j sehingga keseluruhan dataset hanya ada di satu server.
2. Tidak disediakannya keamanan pada tingkat data dan tidak adanya enkripsi data.
3. Tidak adanya audit keamanan di Neo4j [10].

B. Deteksi Komunitas

Sebuah komunitas merupakan sekumpulan node yang mirip satu sama lain, contohnya seperti kelompok orang dengan hobi yang mirip [11]. Struktur komunitas, yang teridentifikasi berdasarkan keberadaan kelompok node yang memiliki hubungan erat dan relatif terpisah dibanding bagian lainnya, menjadi salah satu permasalahan paling menantang dalam analisis jaringan. Menelaah struktur komunitas dapat menghasilkan relasi potensial antar elemen dan membantu dalam pemahaman jaringan secara efektif [12]. Deteksi komunitas adalah cara untuk mengidentifikasi struktur komunitas yang ada di dalam graf dengan memanfaatkan informasi topologi dari graf, yang dianggap sebagai salah satu kegiatan penambangan graf yan krusial [13]. Algoritma deteksi komunitas digunakan untuk mengevaluasi bagaimana kelompok node dikelompokkan atau dipisahkan, serta kecenderungan untuk menguat atau terpecah [3].

C. *Label Propagation*

Algoritma label propagation adalah algoritma cepat untuk menemukan komunitas pada graf. Algoritma ini mendeteksi komunitas menggunakan struktur jaringan sebagai panduan, dan tidak memerlukan fungsi tujuan yang telah ditetapkan atau informasi awal tentang komunitas. Algoritma label propagation bekerja dengan menyebarkan label ke seluruh jaringan dan membentuk komunitas berdasarkan proses penyebaran label yang telah dilakukan. Intuisi di balik algoritma ini adalah satu label dapat dengan cepat menjadi dominan dalam kelompok node yang terhubung secara padat, tetapi akan mengalami kesulitan melintasi wilayah yang jarang terhubung. Label akan terjebak di dalam kelompok node yang terhubung secara padat, dan node yang berakhir dengan label yang sama ketika algoritma selesai dapat dianggap sebagai bagian dari komunitas yang sama [6].

Label propagation merupakan algoritma dengan pendekatan semi-supervised. Performa yang ditunjukkan dari pendekatan supervised memang lebih baik dari unsupervised, meskipun begitu proses pemberian label dalam sampel jumlah besar bisa memakan waktu dan tenaga. Sementara pendekatan semi-supervised menggabungkan antara data dengan label dan yang tidak berlabel [14].

III. METODOLOGI

A. Persiapan Data

Dataset yang digunakan adalah dataset taksonomi laba-laba yang bersumber dari website Kaggle [15]. Dataset diunduh dalam format CSV, sampel dataset yang digunakan digambarkan seperti pada tabel 1.

TABEL I
SAMPEL DATASET

species id	species_lsid	family	genus	species	subspecies	author	year	parentheses	distribution
1	urn:lsid:nmbe.ch:spidersp:000896	Actinopodidae	Actinopus	caraiba		Simon	1889	1	Venezuela
2	urn:lsid:nmbe.ch:spidersp:000898	Actinopodidae	Actinopus	crassipes		Keyserling	1891	1	Brazil, Paraguay, Argentina
3	urn:lsid:nmbe.ch:spidersp:000899	Actinopodidae	Actinopus	cucutaensis		Mello-LeitÃ£o	1941	0	Colombia, Venezuela, Brazil
4	urn:lsid:nmbe.ch:spidersp:000900	Actinopodidae	Actinopus	dubiomaculatus		Mello-LeitÃ£o	1923	0	Brazil
5	urn:lsid:nmbe.ch:spidersp:000901	Actinopodidae	Actinopus	echinus		Mello-LeitÃ£o	1949	0	Brazil

Dataset terdiri dari 49738 baris dengan 10 kolom yaitu:

1. Species_Id
   Kolom ini menunjukkan id unik berupa integer increment.
2. Species_lsid
   Kolom ini merupakan id yang dapat digunakan untuk menamai mencari lebih lanjut tentang spesies melalui internet.
3. Family
   Keluarga ilmiah yang dimiliki oleh spesies tersebut
4. Genus
   Marga, yaitu kategori taksonomi yang berada di bawah keluarga dan di atas spesies.
5. Species
   Nama spesies dari laba-laba.
6. Subspecies
   Nama subspesies dari laba-laba.
7. Author
   Nama penemu dari spesies laba-laba.
8. Year
   Tahun spesies itu ditemukan.
9. Parentheses
   Saat peneliti tidak yakin apakah suatu populasi merupakan subspesies atau spesies lengkap, nama spesies dapat ditulis dalam tanda kurung. Bernilai 0 jika tidak terdapat tanda kurung dan bernilai1 jika terdapat tanda kurung.
10. Distribution
    Negara/wilayah spesies laba-laba ditemukan.

Dalam penelitian ini dilakukan praproses data agar pemrosesan algoritma dapat dilakukan lebih efisien sesuai tujuan penelitian. Untuk itu kolom species_Id, species_lsid, family, genus, author, subspecies dan parentheses akan dihilangkan karena tidak akan digunakan dalam analisis deteksi komunitas.

Setelah dilakukan praproses data, dataset kemudian dimuat dalam Neo4j. Dataset diunggah ke platform Github lalu dipanggil dalam kode Neo4j. Kode yang digunakan untuk melakukan pemuatan dataset adalah:

```
LOAD CSV WITH HEADERS FROM
'https://raw.githubusercontent.com/imammansyur/dataset/main/species_export_20211111.csv'
AS row  WITH row LIMIT 1000
MERGE (s:Species {name: row.species})
WITH s, row
UNWIND split(row.distribution, ', ') as distribution
MERGE (d:Distribution {name: distribution})
MERGE (s)-[dd:DITEMUKAN_DI]->(d)
```

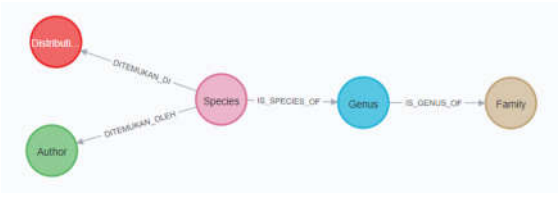
Pada dataset yang dimuat dibuat 2 node yaitu Species dan Distribution. Node Species menunjukkan nama spesies laba-laba, sementara node Distribution menunjukkan nama negara. Kedua node tersebut dihubungkan oleh relasi DITEMUKAN_DI yang menunjukkan bahwa suatu spesies ditemukan di negara tersebut.

Pada kolom distribution di dataset mentah, apabila suatu spesies ditemukan di lebih dari satu negara maka dituliskan dengan dipisahkan tanda koma. Karena itu perlu dilakukan pemisahan baris untuk masing-masing negara dari kolom tersebut pada saat dataset dimuat dalam Neo4j. Jumlah node, properties dan relationship pada dataset digambarkan pada tabel 2.

TABEL II
JUMLAH NODE, PROPERTY, DAN RELATIONSHIP DARI DATASET

Nodes	Properties	Relationship
1385	3333	3175

B. Pemodelan graf



Gbr 1 Skema dataset

Gbr 2 Skema data yang dipakai

Dataset yang dibuat memiliki skema seperti pada gambar 2. Untuk pembuatan algoritma deteksi komunitas spesies laba-laba dan negara habitat, dimodelkan graf bernama "speciesDistribution" yang terdiri dari node species yang dihubungkan ke node distribution dengan relasi DITEMUKAN_DI dengan skema seperti pada gambar 3. Berikut merupakan kode yang digunakan untuk memodelkan graf relasi DITEMUKAN_DI antara node species ke node distribution.

```
CALL gds.graph.create(
  'speciesDistribution',
  ['Species', 'Distribution'],
  'DITEMUKAN_DI'
)
YIELD
  graphName AS graph,
  relationshipProjection AS
deskripsiRelationship,
  nodeCount AS nodes,
  relationshipCount AS relCount
```

Graf speciesDistribution yang dibuat memiliki node berjumlah 1133 dan relasi berjumlah 1157. Algoritma label propagation akan dijalankan pada graf speciesDistribution untuk mendeteksi komunitas spesies laba-laba.

C. Pemrosesan algoritma

```
CALL
gds.labelPropagation.stream('speciesDistribution')
YIELD nodeId, communityId AS Community
RETURN gds.util.asNode(nodeId).name AS Name, Community
ORDER BY Community, Name
```

Selanjutnya algoritma label propagation dijalankan pada graf speciesDistribution. Algoritma ini memberikan keluaran berupa kolom nama node dan nomor komunitas dimana spesies tergabung.

IV. HASIL DAN PEMBAHASAN

A. Hasil pemrosesan algoritma

Berikut merupakan sampel 10 baris pertama keluaran dari algoritma label propagation yang telah dijalankan pada graf speciesDistribution. Tabel 3 menunjukkan bahwa negara Venezuela; spesies caraiba, cucutaensis, rojasi, scalops dan valencianus berada dalam satu komunitas dengan nomor 6. Sementara negara Brazil; spesies crassipes, dubiomaculatus, dan echinus berada dalam satu komunitas dengan nomor 7.

TABEL III
HASIL DATA 10 BARIS PERTAMA

No	Name	Community
1	Venezuela	6
2	caraiba	6
3	cucutaensis	6
4	rojasi	6
5	scalops	6
6	valencianus	6
7	Brazil	7
8	crassipes	7
9	dubiomaculatus	7
10	echinus	7

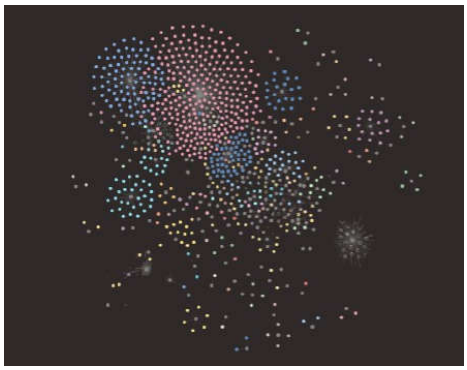
Kemudian untuk mempermudah pemahaman hasil deteksi komunitas dalam skala yang lebih luas, hasil deteksi komunitas akan ditampilkan dalam bentuk graf. Untuk itu hasil algoritma

```
CALL
gds.labelPropagation.write('speciesDistribution
', { writeProperty: 'community' })
YIELD communityCount, ranIterations,
didConverge
```

perlu disimpan dalam properti node species dan distribution. Berikut adalah kode yang digunakan untuk menyimpan hasil algoritma ke dalam properti yang bernama community.

B. Visualisasi graf hasil algoritma

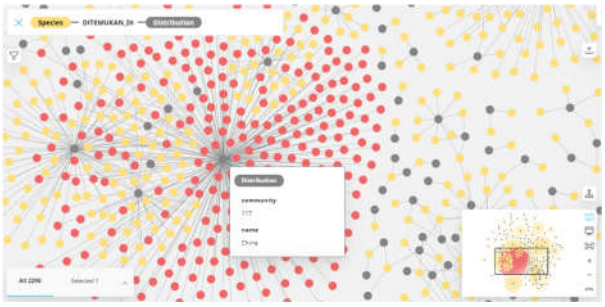
Hasil dari deteksi komunitas spesies laba-laba menggunakan label propagation menunjukkan bahwa terdapat 159 komunitas yang terbentuk. Visualisasi graf dari hasil deteksi komunitas spesies laba-laba dengan menggunakan Neo4j Bloom terdapat pada gambar 4.



Gbr 3 Visualisasi graf komunitas laba-laba

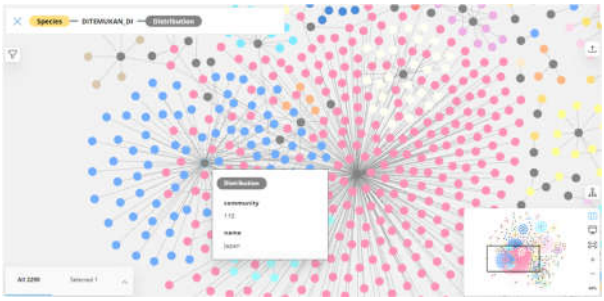
Pada gambar 4 terdapat satu komunitas besar yang berwarna merah muda, jika ditelusuri lebih detil komunitas tersebut adalah komunitas negara China dengan nomor 117. Proporsi komunitas negara China terhadap keseluruhan data dapat dilihat pada peta mini di pojok kanan bawah pada gambar 5, warna merah menunjukkan komunitas negara China.

```
MATCH (s:Species)-
[:DITEMUKAN_DI]-(d:Distribution)
RETURN d.name as nama_negara,
d.community as no_komunitas,
COUNT(s.community) as anggota
ORDER BY anggota desc
```

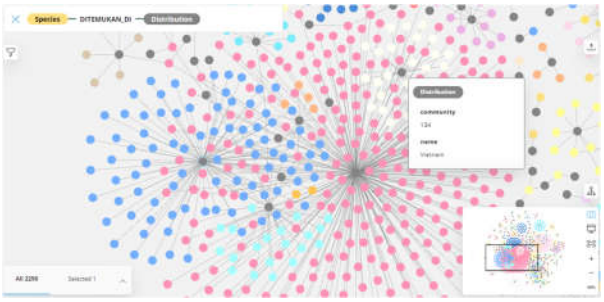


Gbr 4 Komunitas negara China dan perbandingannya terhadap komunitas spesies lainnya

Hal menarik lainnya adalah di sekitar komunitas negara China (merah muda) juga terdapat komunitas lain yang lebih kecil yaitu Jepang (Biru), Vietnam (Putih) dan Korea (Biru muda). Ketiga komunitas tersebut terletak di dalam daerah komunitas negara China yang ditampilkan pada gambar 6, gambar 7 dan gambar 8.



Gbr 5 Perbandingan komunitas China terhadap Jepang



Gbr 6 Perbandingan komunitas China terhadap Vietnam

Gbr 7 Perbandingan komunitas China terhadap Korea

Sayangnya visualisasi graf yang ditampilkan belum cukup menggambarkan ukuran dan identitas dari komunitas, karena itu dilakukan kueri untuk menghitung ukuran dari masing-masing komunitas sebagai berikut beserta sampel hasilnya yaitu 10 negara dengan komunitas spesies laba-laba terbesar yang ditampilkan pada table 4.

TABEL IV
KOMUNITAS BESERTA JUMLAH ANGGOTANYA

"nama_negara"	"no_komunitas"	"anggota"
"China"	117	279
"Japan"	110	86
"USA"	235	86
"New Zealand"	290	83
"Nepal"	186	49
"Korea"	128	32
"Thailand"	93	29
"Vietnam"	134	28
"Japan (Ryukyu Is.)"	423	21
"Italy"	344	20

V. KESIMPULAN

Berdasarkan hasil penelitian dapat disimpulkan bahwa algoritma label propagation dapat mengelompokkan komunitas spesies laba-laba sebanyak 159 komunitas. Negara China merupakan negara dengan komunitas spesies laba-laba terbesar di dunia dengan jumlah spesies laba-laba berjumlah 279 spesies, diikuti oleh negara Jepang, Amerika Serikat, Selandia Baru, Nepal, Korea, Thailand, Vietnam, Jepang (Pulau Ryukyu), dan Italia sebagai 10 besar negara dengan komunitas spesies laba-laba terbesar.

REFERENSI

[1] Z. Q. Zhang, "Phylum arthropoda," *Zootaxa*, vol. 3703, no. 1, 2013, doi: 10.11646/zootaxa.3703.1.6.

[2] W. Hu, Z. Mei, Y. Liu, Z. Yu, F. Zhang, and M. Duan, "Recovered Grassland Area rather than Plantation Forest Could Contribute More to Protect Epigeic Spider Diversity in Northern China," *Agric. Ecosyst. Environ.*, p. 107726, Nov. 2021, doi: 10.1016/J.AGEE.2021.107726.

[3] Neo4j, "Community Detection - Neo4j Graph Data Science," *Neo4j*, 2021. https://neo4j.com/docs/graph-data-science/current/algorithms/community/ (accessed Dec. 17, 2021).

[4] J. Golbeck, *Analyzing the Social Web*. 2013.

[5] P. N. Indrianingtyas, F. A. Formen, and N. A. Rakhmawati, "ANALISIS SUKU DI INDONESIA MENGGUNAKAN ALGORITMA CLOSENESS CENTRALITY," *POSITIF J. Sist. dan Teknol. Inf.*, vol. 6, no. 2, pp. 137–142, Dec. 2020, doi: 10.31961/POSITIF.V6I2.923.

[6] Neo4j, "Label Propagation - Neo4j Graph Data Science," *Neo4j*, 2021. https://neo4j.com/docs/graph-data-science/current/algorithms/label-propagation/ (accessed Dec. 17, 2021).

[7] L. P. G. Widiastuti, N. I. Maghfirani, and N. A. Rakhmawati, "Klasterisasi Pahlawan Nasional Indonesia Berdasarkan Daerah Asal Menggunakan Algoritma Community Detection," *SISTEMASI*, vol. 10, no. 1, 2021, doi: 10.32520/stmsi.v10i1.1068.

[8] N. A. Rakhmawati and K. Mufidah, "Social Network Analysis of Legislative Candidates in Indonesia General Election 2019 Using Community Detection," 2020, doi: 10.1109/IC2IE50715.2020.9274669.

[9] Z. Zhu, X. Zhou, and K. Shao, "A novel approach based on Neo4j for multi-constrained flexible job shop scheduling problem," *Comput. Ind. Eng.*, vol. 130, 2019, doi: 10.1016/j.cie.2019.03.022.

[10] N. Gupta and R. Agrawal, "NoSQL Security," in *Advances in Computers*, vol. 109, 2018.

[11] K. Wu and X. Liu, "Community detection in directed acyclic graphs of adversary interactions," *Phys. A Stat. Mech. its Appl.*, vol. 584, 2021, doi: 10.1016/j.physa.2021.126370.

[12] B. Yang, X. Huang, W. Cheng, T. Huang, and X. Li, "Discrete bacterial foraging optimization for community detection in networks," *Futur. Gener. Comput. Syst.*, vol. 128, 2022, doi: 10.1016/j.future.2021.10.015.

[13] Y. Kang, J.-S. Lee, W.-Y. Shin, and S.-W. Kim, "Community Reinforcement: An Effective and Efficient Preprocessing Method for Accurate Community Detection," *Knowledge-Based Syst.*, p. 107741, Nov. 2021, doi: 10.1016/J.KNOSYS.2021.107741.

[14] F. Wang, L. Zhu, L. Xie, Z. Zhang, and M. Zhong, "Label propagation with structured graph learning for semi-supervised dimension reduction[Formula presented]," *Knowledge-Based Syst.*, vol. 225, 2021, doi: 10.1016/j.knosys.2021.107130.

[15] I. M. Solichin, B. N. Jati, and N. A. Rakhmawati, "Dataset Spesies Laba-Laba," Dec. 2021, doi: 10.5281/ZENODO.5791674.