Departemen
Ilmu Komputer
http://cs.ipb.ac.id/

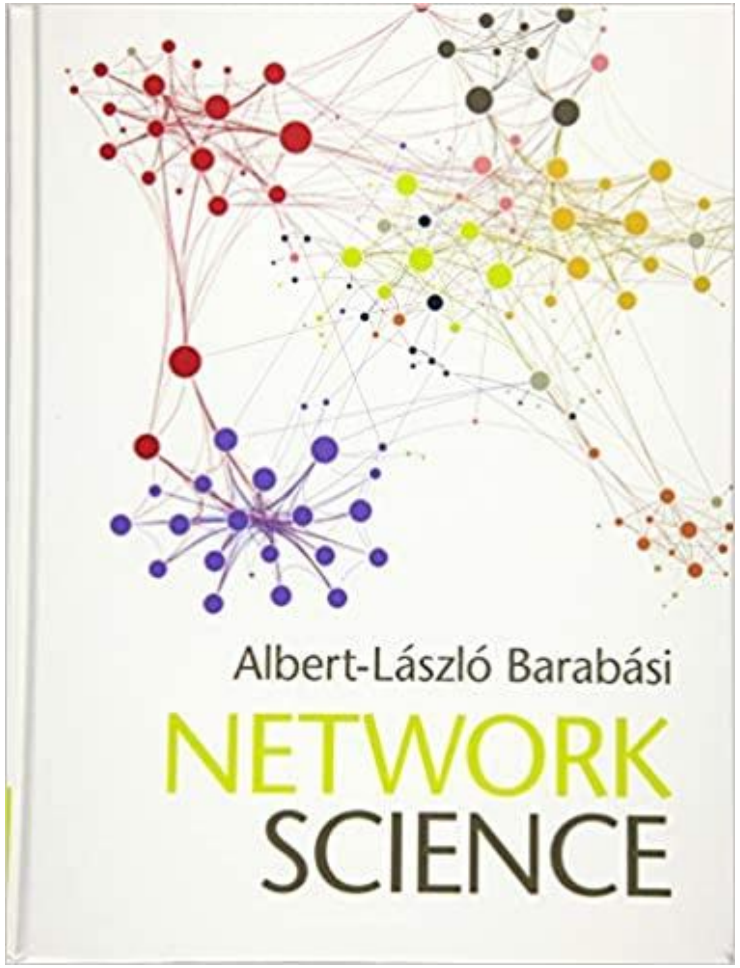
A Network Analysis to Identify Lung Cancer Comorbid Diseases

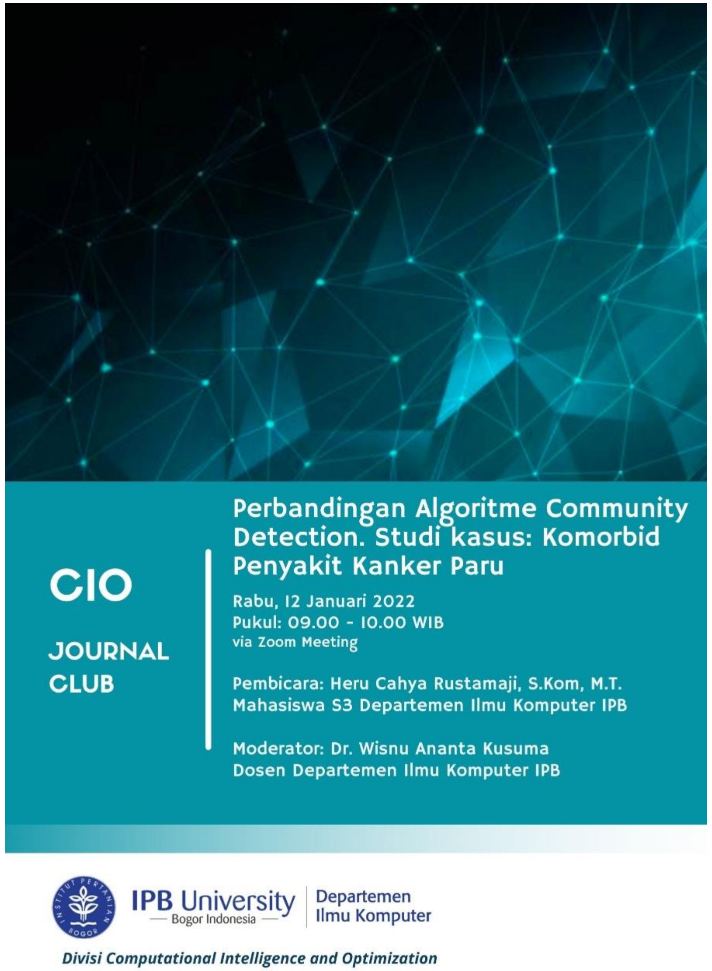
Heru Cahya Rustamaji, Yustina S. Suharini, Angga A Permana
Wisnu Ananta Kusuma, Sri Nurdiati, Irmanida Batubara

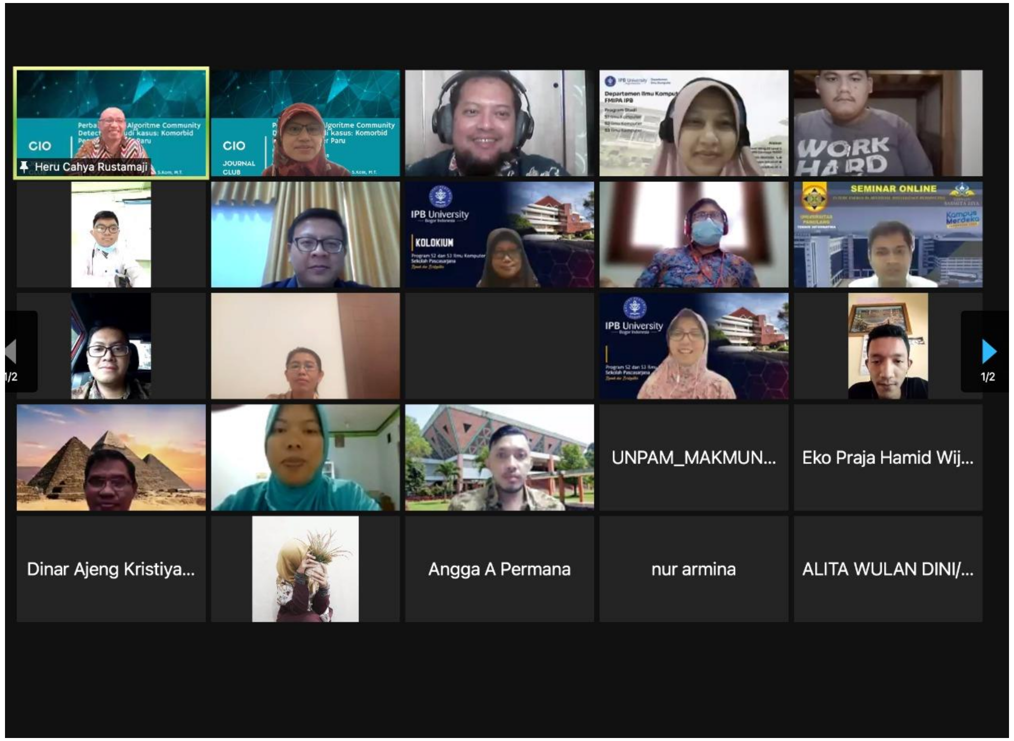
Dosen Ilmu Komputer, FMIPA, IPB
Peneliti Pusat Studi Biofarmaka Tropika, LPPM, IPB (PUI-PT)
ananta@apps.ipb.ac.id, 081280983486 (WA)

Weekly Seminar, 5 Agustus 2022, 16.00-17.00
Indonesian Association for Pattern Recognition

KLUB BUKU DAN JURNAL KELOMPOK RISET BIOINFORMATIKA ILMU KOMPUTER, FMIPA, IPB







Setiap Jum’at pukul 16.00

Rustamaji *et al. Applied Network Science* (2022) 7:30
https://doi.org/10.1007/s41109-022-00466-y

Applied Network Science

RESEARCH **Open Access**

A network analysis to identify lung cancer comorbid diseases

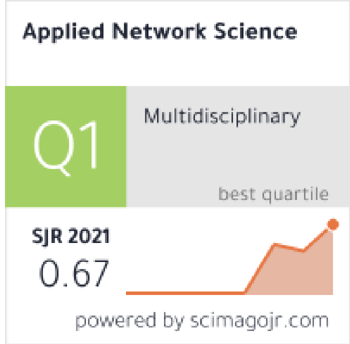
Heru C. Rustamaji[1,6], Yustina S. Suharini[1,7], Angga A. Permana[1,8], Wisnu A. Kusuma[1,4*], Sri Nurdiati[2], Irmanida Batubara[3,4] and Taufik Djatna[5]

*Correspondence:
ananta@apps.ipb.ac.id

[1] Department of Computer Science, Faculty of Mathematics and Natural Sciences, IPB University, Bogor, Indonesia
Full list of author information is available at the end of the article

Abstract

Cancer patients with comorbidities face various life problems, health costs, and quality of life. Therefore, determining comorbid diseases would significantly affect the treatment of cancer patients. Because cancer disease is very complex, we can represent the relationship between cancer and its comorbidities as a network. Furthermore, the network analysis can be employed to determine comorbidities as a community detection problem because the relationship between cancer and its comorbidities forms a community. This study investigates which community detection algorithms are more appropriate to determine the comorbid of cancer. Given different community findings, this study attempted to analyze the modularity generated by the algorithm to decide the significant comorbid diseases. We retrieved lung cancer comorbid data on the



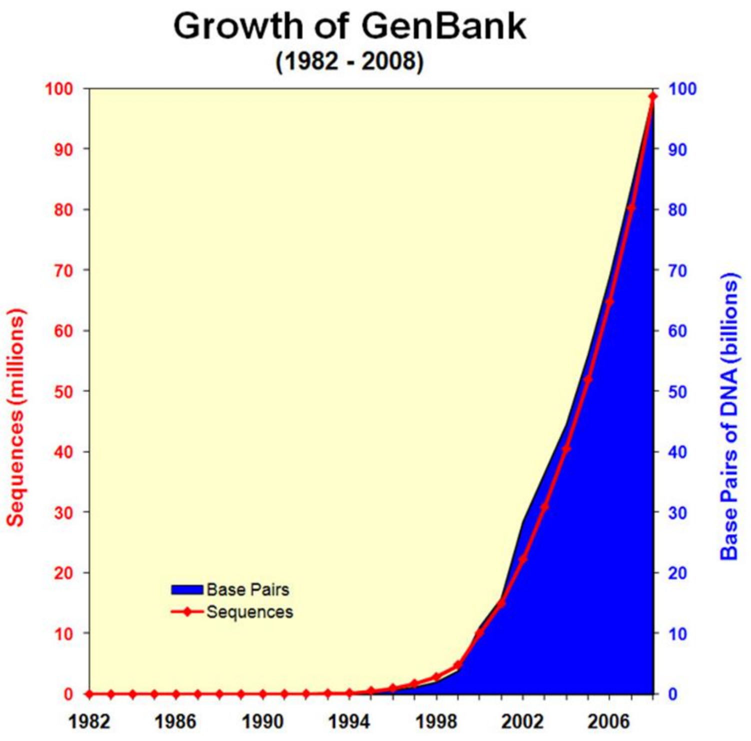
Outline

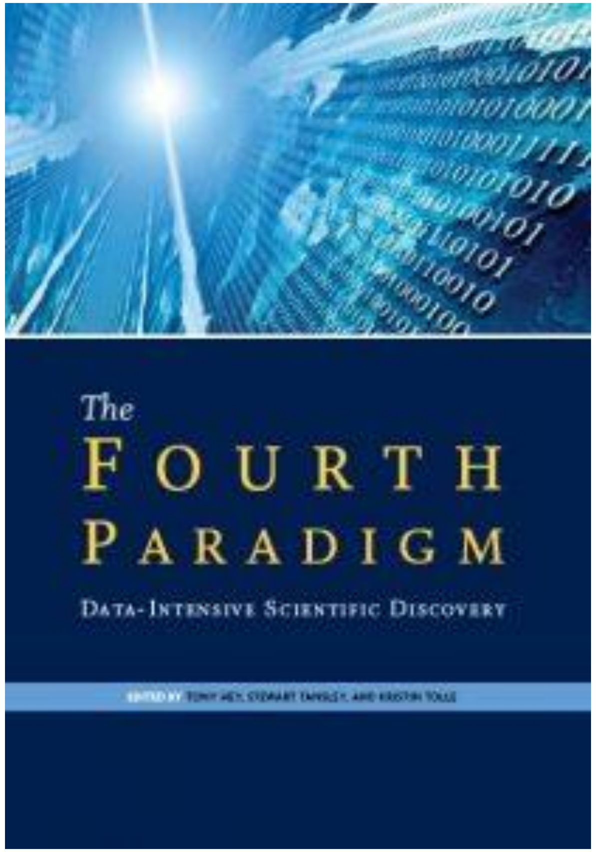
- Motivasi
- Komunitas dan Komorbiditas
- Deteksi Komunitas
- Material, Metode, dan Hasil
- Kesimpulan

MOTIVASI

Data-Intensive Scientific Discovery

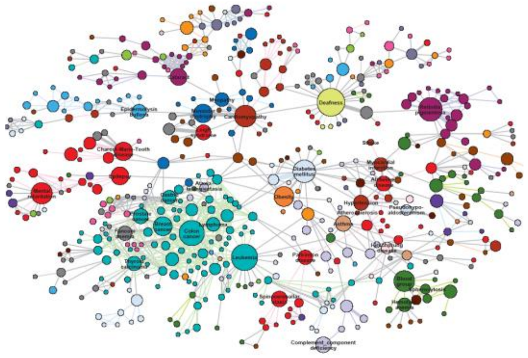
- Pergeseran dari paradigma komputasi ke paradigma data
- Terobosan saintifik dan komputasi tingkat tinggi memacu **eksplorasi data**



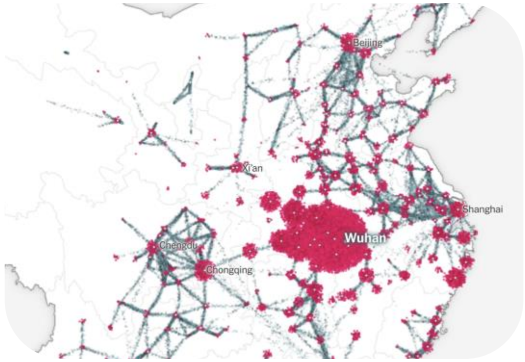


Tony Hey, Stewart Tansley, Kristin Tolle
Published by Microsoft Research, 2009

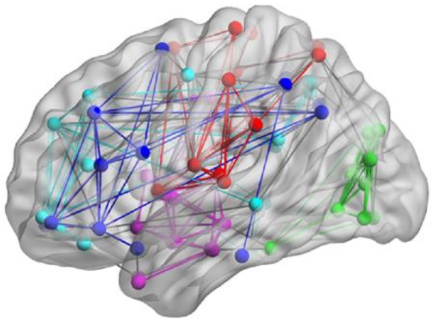
Jejaring dan Analisisnya



Jejaring penyakit



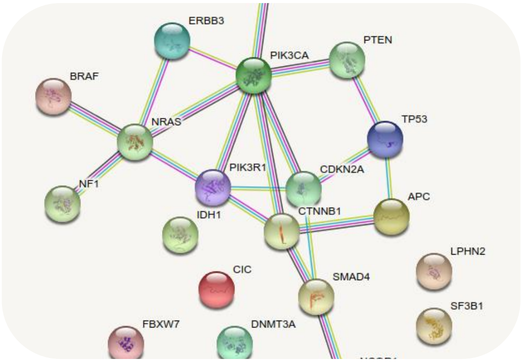
Jejaring penyebaran penyakit Covid



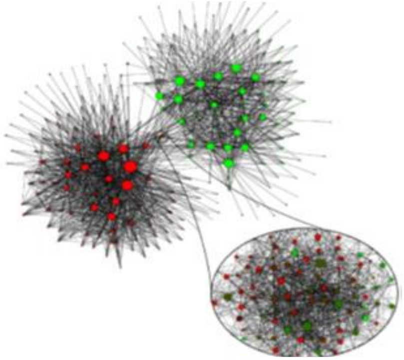
Human Brain

(10 – 100 juta neuron)

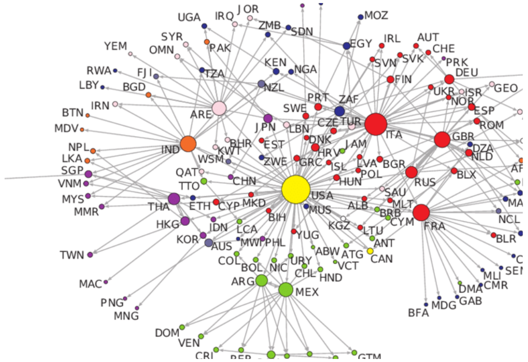
social network



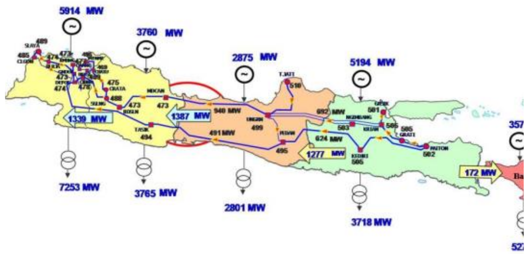
Protein Interaction Network



Communication network



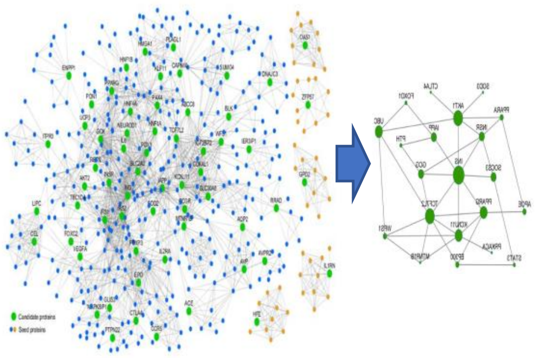
Trade Network



Power Grid

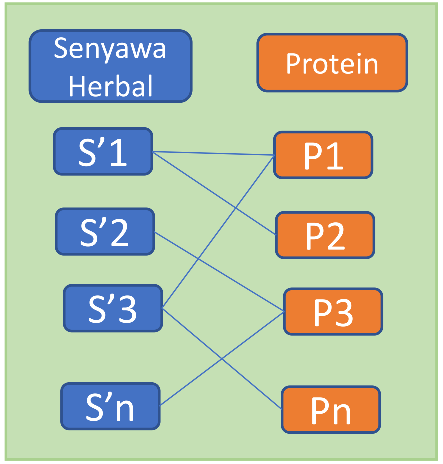
Contoh Jejaring dan Analisisnya: ijah.apps.cs.ipb.ac.id

Menemukan protein yang memiliki peran penting terkait penyakit



- Topologi
- Skyline
- Klustering (Markov, ClusterOne, fuzzy)

Memprediksi hubungan/interaksi dari jejaring senyawa-protein

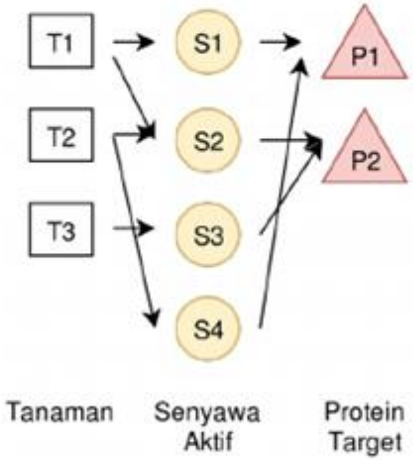


Link prediction
- BLM-NII
- Deep learning

Similarity measure
- GA

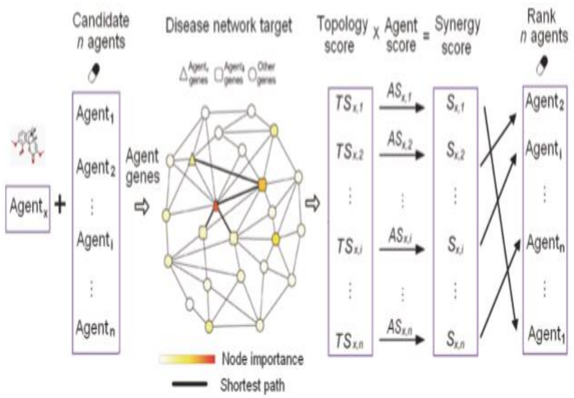
Imbalance data
- SMOTE, Deep Learning

Melakukan penelurusan graf untuk menemukan formula jamu



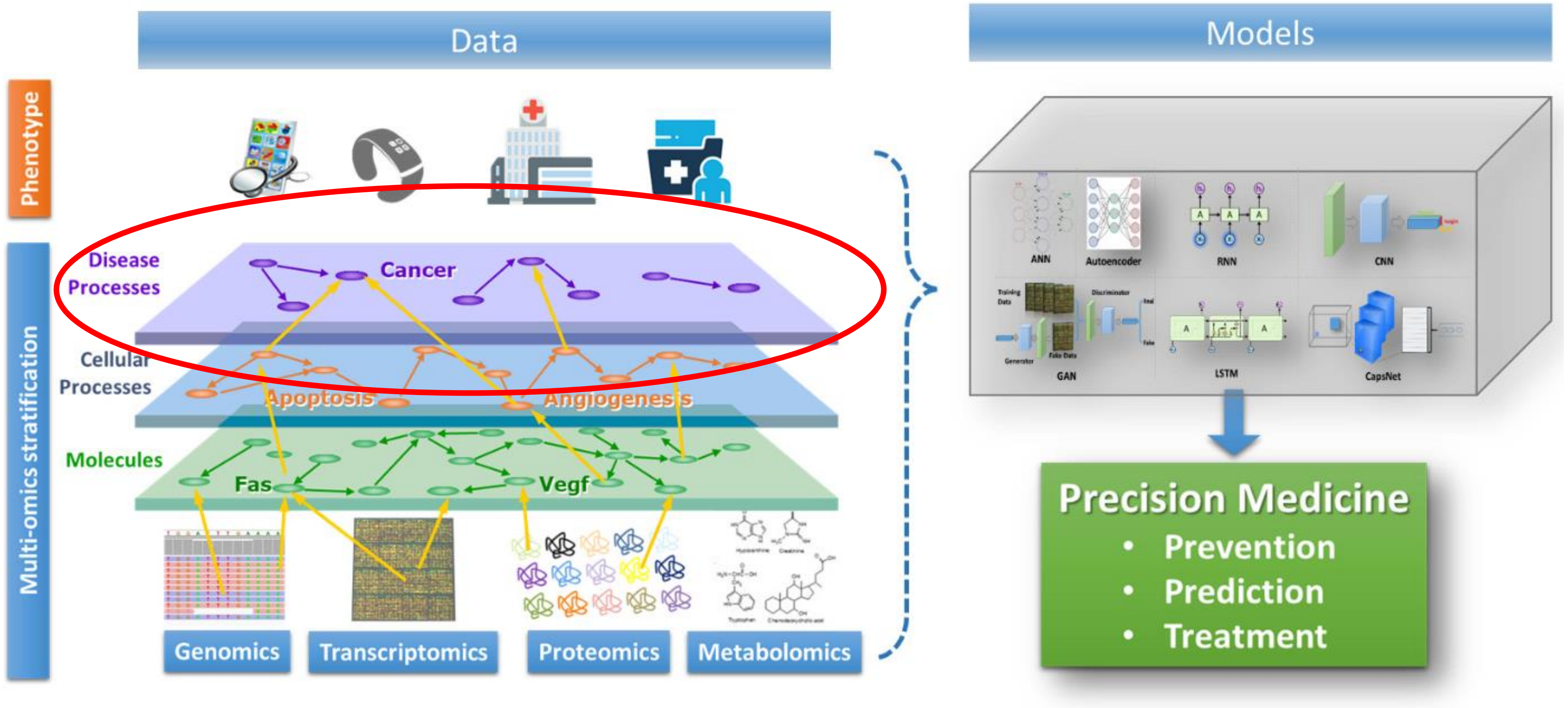
- Complete Search
- Branch and Bound
- Complete Search + GPU

Menghitung skor sinergi



- Sinergi dua senyawa (NIMS)
- Sinergi tiga senyawa

MODEL OMICS UNTUK PENGEMBANGAN PRECISION MEDICINE

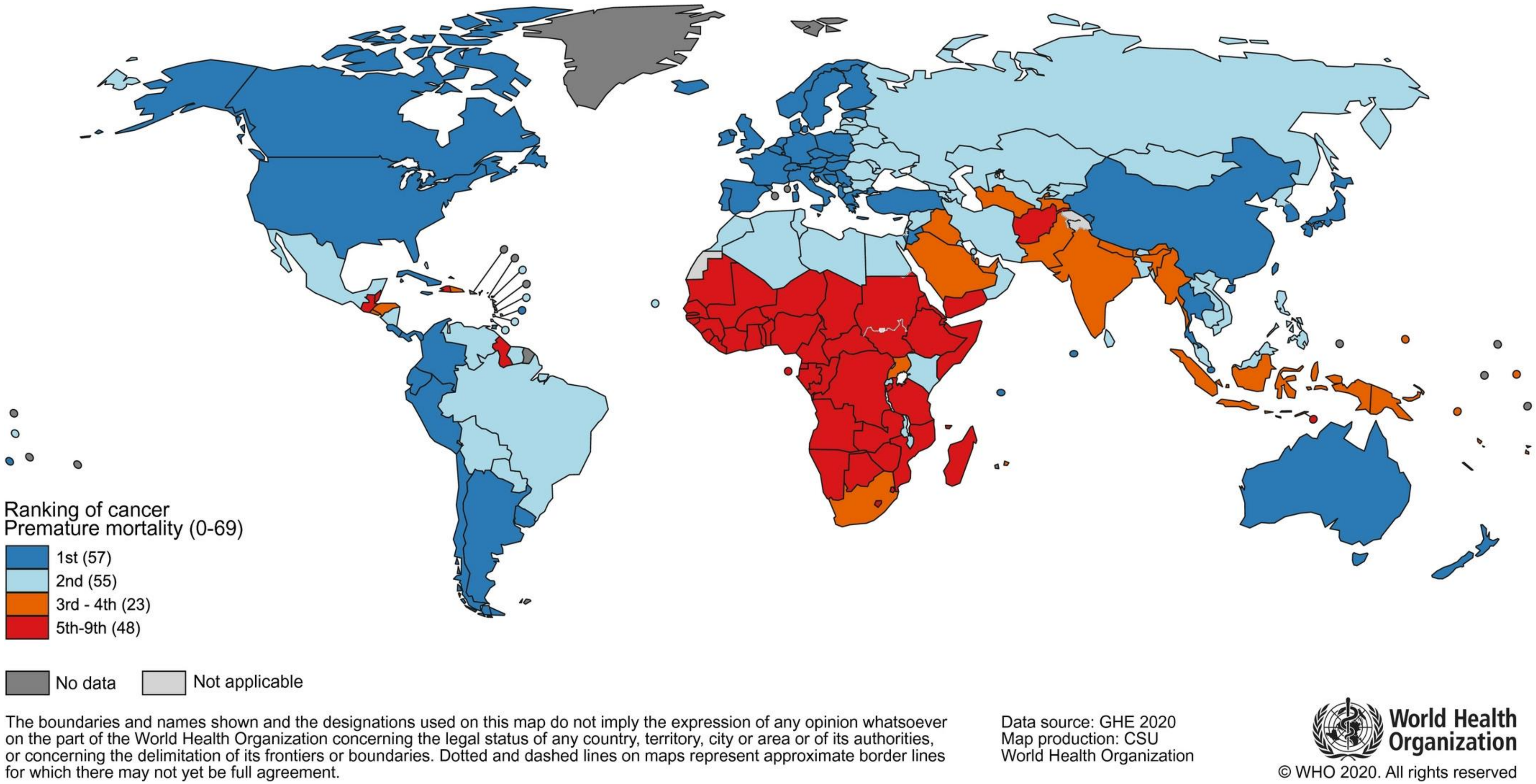


L. Koumakis, Deep learning models in genomics; are we there yet?, Computational and Structural Biotechnology Journal 18 (2020) 1466–1473

KANKER

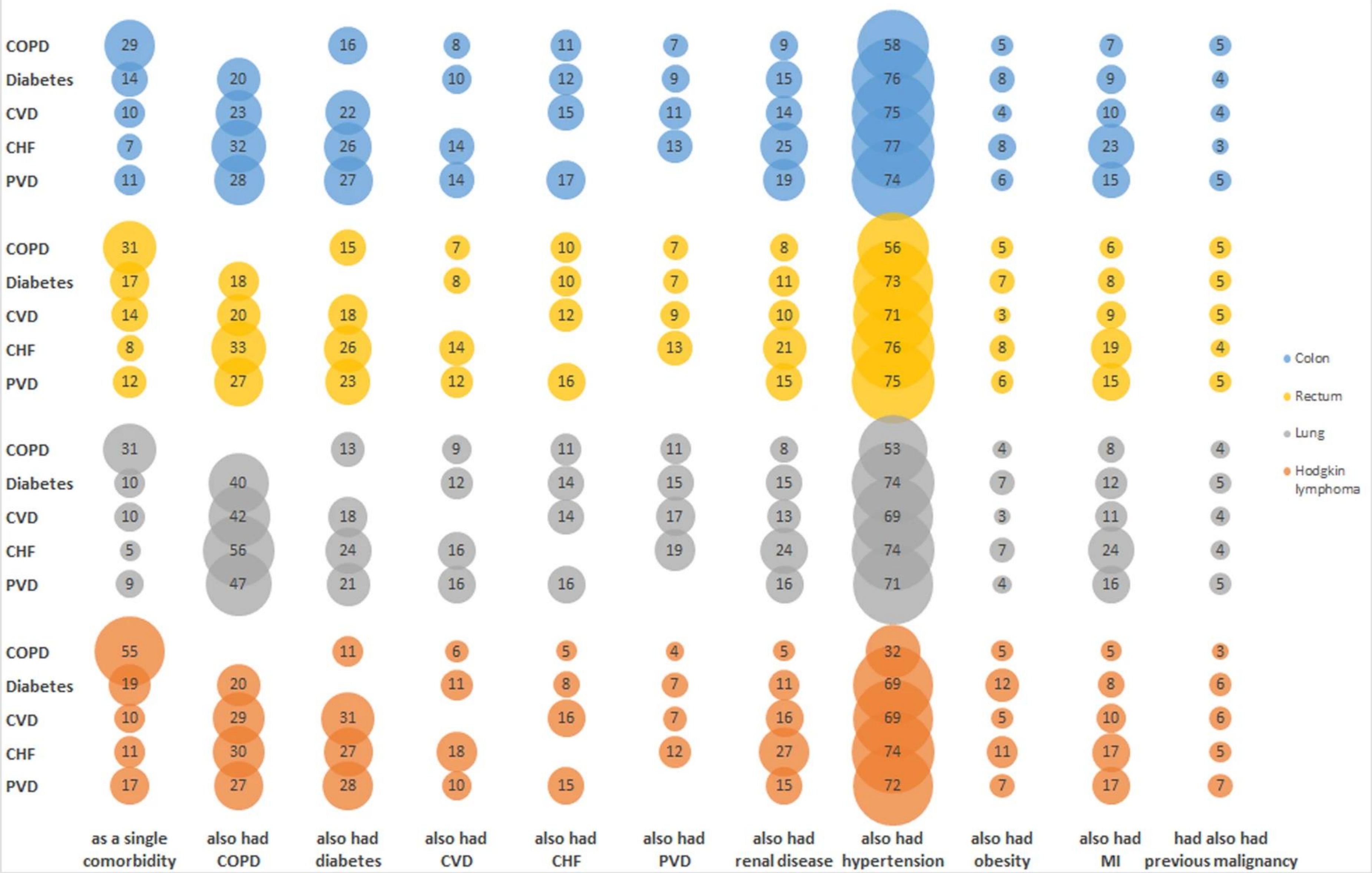
Penyakit yang ditandai dengan pertumbuhan yang tidak terkendali dan penyebaran sel abnormal yang dapat mengakibatkan kematian

Kanker menjadi salah satu penyebab kematian utama sebelum usia 70 tahun di 112 negara pada 2019

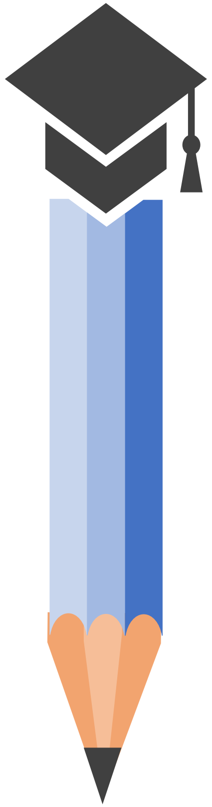


H. Sung *et al.*, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA. Cancer J. Clin.*, p. caac.21660, Feb. 2021, doi: 10.3322/caac.21660.

KANKER PARU DAN KOMORBIDITASNYA

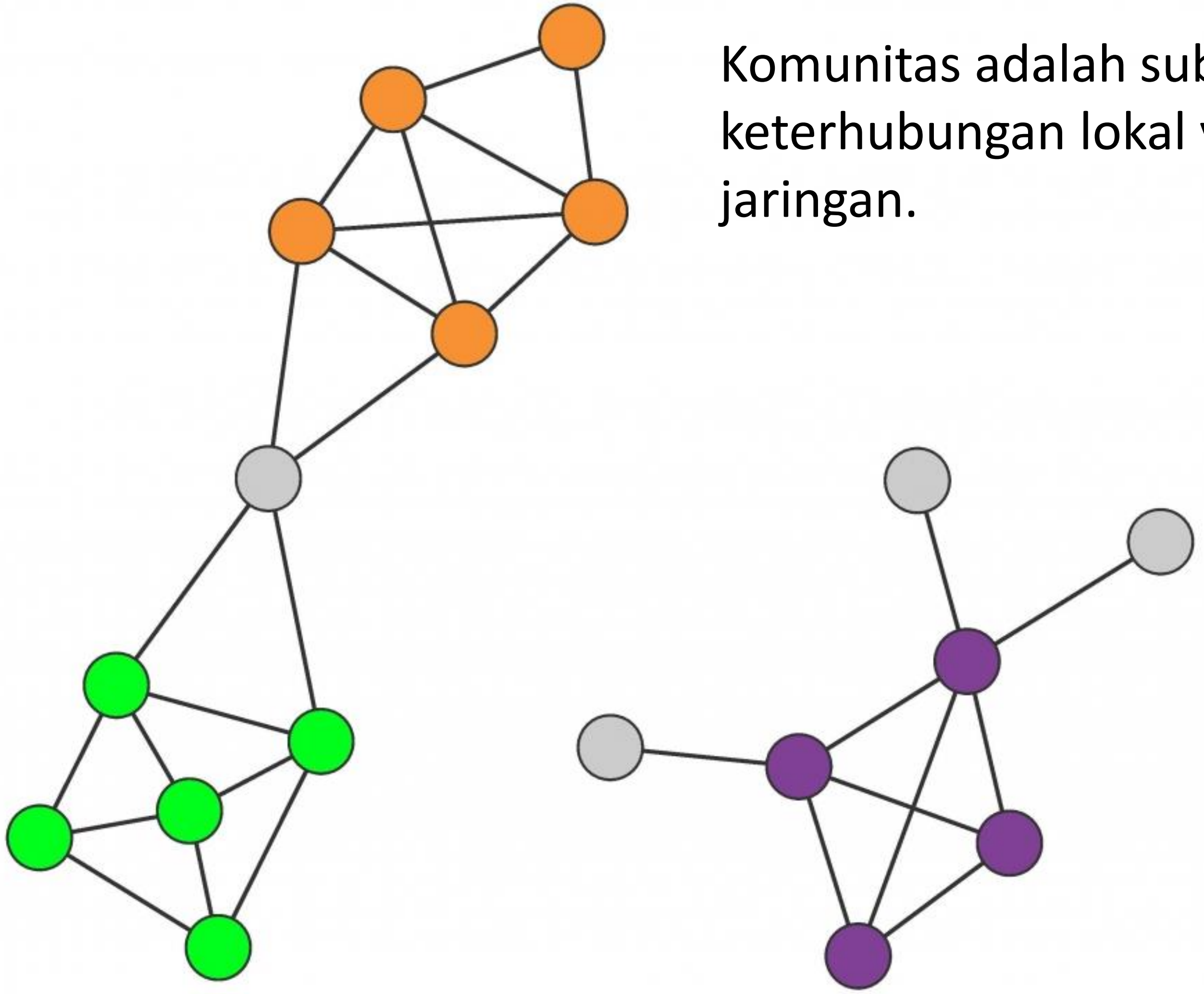


H. Fowler *et al.*, “Comorbidity prevalence among cancer patients: A population-based cohort study of four cancers,” *BMC Cancer*, vol. 20, no. 1, pp. 1–15, Jan. 2020, doi: 10.1186/s12885-019-6472-9.



KOMUNITAS DAN KOMORBIDITAS

COMMUNITY (KOMUNITAS)



Komunitas adalah subgraf yang memiliki keterhubungan lokal yang padat dalam suatu jaringan.

Hipotesis keterhubungan:

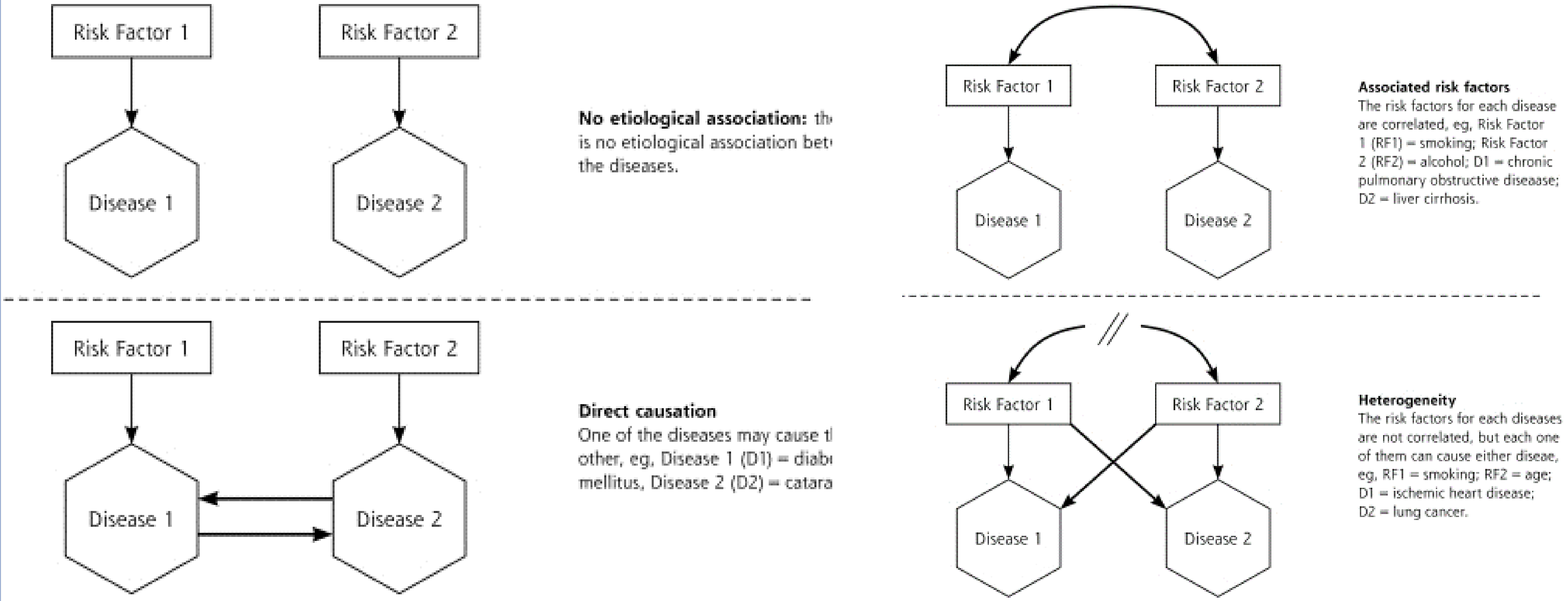
- Setiap komunitas berkorespondensi dengan sebuah subgraf yang memiliki keterhubungan lokal
- Jika sebuah network memiliki dua komponen yang terisolasi, maka setiap komunitas hanya menjadi bagian dari satu komponen
- Pada komponen yang sama, suatu komunitas tidak memiliki dua subgraf

Hipotesis Kepadatan:

- Node pada komunitas cenderung untuk terhubung dengan node yang lain pada komunitas yang sama dibandingkan dengan node pada komunitas yang lain

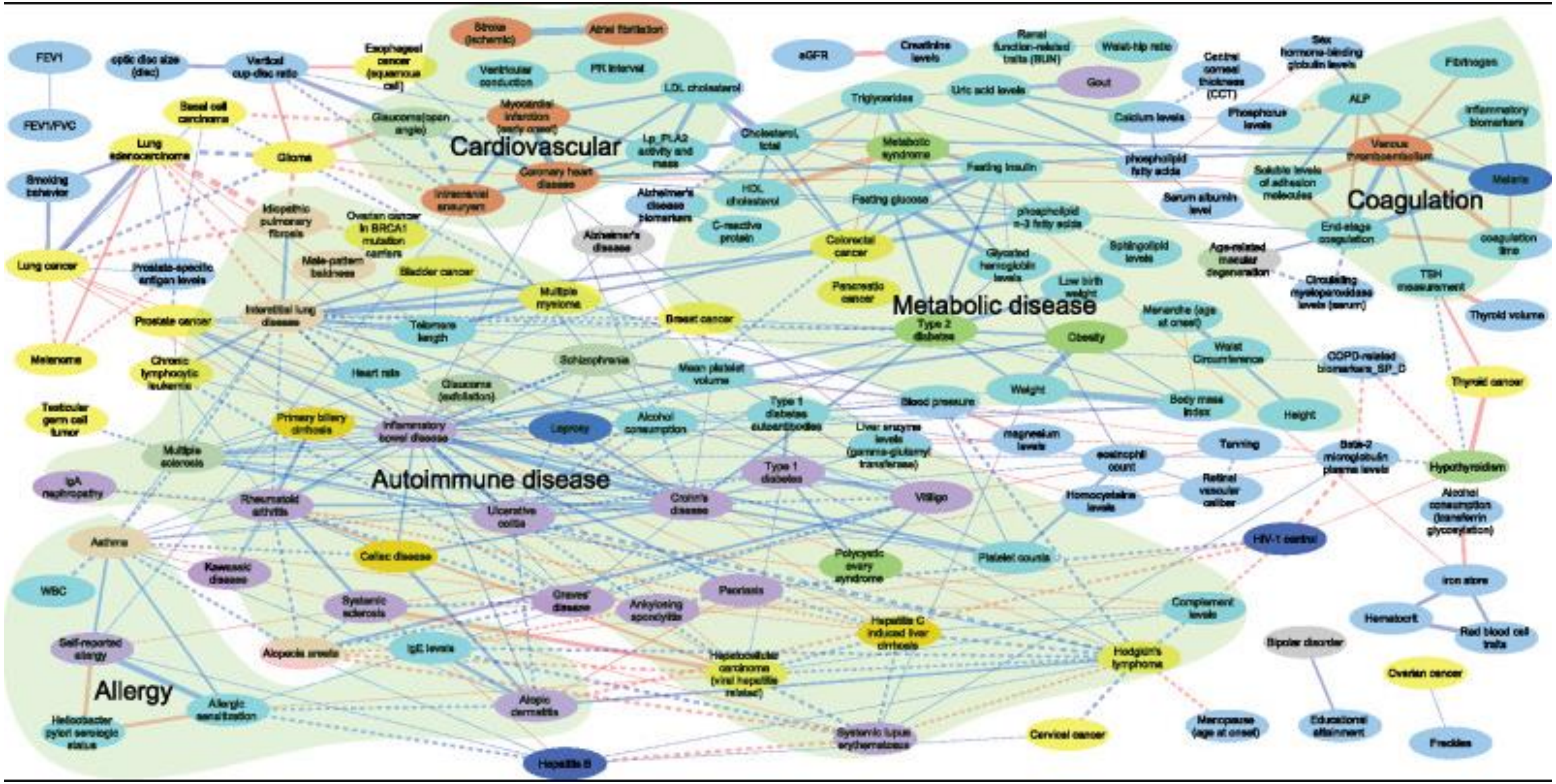
COMMORBIDITY (KOMORBIDITAS)

Comorbidity is associated with worse health outcomes, more complex clinical management, and increased health care costs. There is no agreement, however, on the meaning of the term, and related constructs, such as multimorbidity, morbidity burden, and patient complexity, are not well conceptualized.



Valderas JM, Starfield B, Sibbald B, Salisbury C, Roland M. Defining comorbidity: implications for understanding health and health services. Ann Fam Med. 2009 Jul-Aug;7(4):357-63. doi: 10.1370/afm.983. PMID: 19597174; PMCID: PMC2713155.

KOMUNITAS dan KOMORBIDITAS



Represenatasi jejaring dari 341 asosiasi antara 139 penyakit

J. H. Ohn, “The landscape of genetic susceptibility correlations among diseases and traits,” *J. Am. Med. Informatics Assoc.*, vol. 24, no. 5, pp. 921–926, Sep. 2017, doi: 10.1093/jamia/ocx026

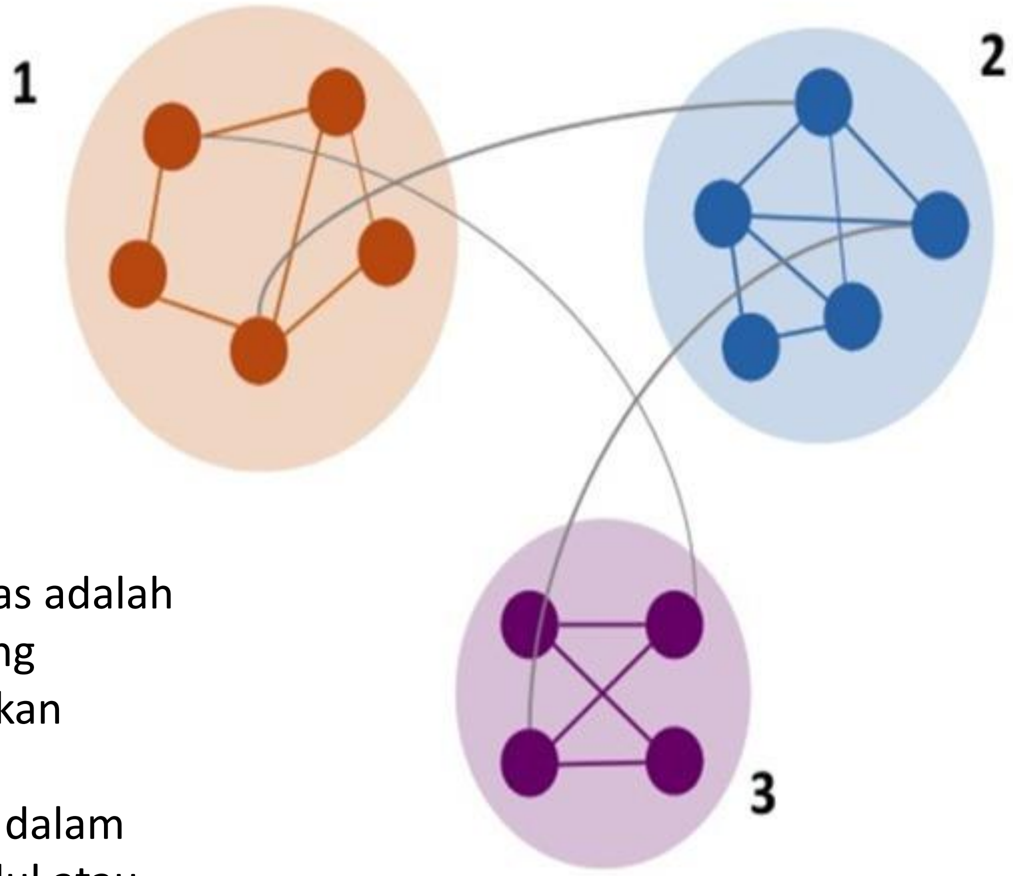
PENELITIAN/STUDI JEJARING KOMORBIDITAS

Table 8 Comparison of comorbidity network studies

Author	Data	Disease	Network constructed	Analysis
Folino et al. (2010)	Medical records of 1462 patients	Common	Occurrence in patient and relative risk	Association rule
Ljubic et al. (2020)	SID California inpatient database (ICD-9)	Colorectal cancer	φ-correlation and Relative Risk (RR)	Centrality measurement
Chmiel et al. (2014)	Database of the Main Association of Austrian Social Security Institutions and Text Mining Pubmed	Common	Statistical multiplex network	Evolution disease network
Renteria-ramos et al. (2018)	Three independent administrative databases Risaralda province (2011–2016)	Common	k-Communities	Intensity analysis and motif coherence
Moratalla-Navarro et al. (2020)	285,342 patients in Catalonia, Spain, (period: 2006–2017)	Hypo thyroidism	Comorbidity networks using logistic regression models	Multivariate logistic regression with LASSO
Our research	Pubtator text mining	Lung Cancer	Disease Similarity	Community Network and Centrality measurement

DETEKSI KOMUNITAS

MODULARITAS



Modularitas adalah ukuran yang menunjukkan kerapatan hubungan dalam suatu modul atau komunitas

$$M_c = \frac{L_c}{L} - \left(\frac{k_c}{2L}\right)^2$$

Community 1: $\left[\frac{6}{20} - \left(\frac{14}{40}\right)^2\right]$ = 0.1775

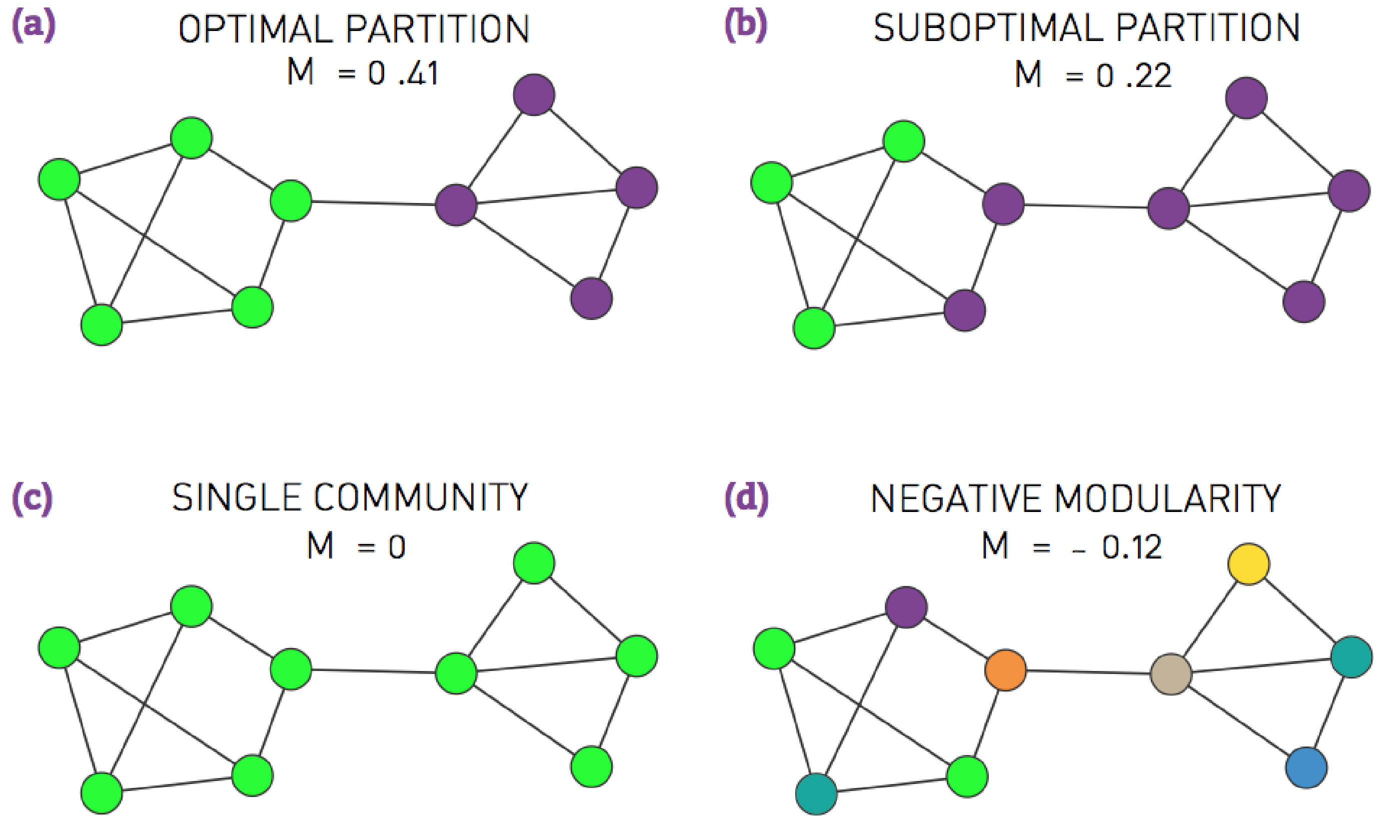
Community 2: $\left[\frac{7}{20} - \left(\frac{16}{40}\right)^2\right]$ = 0.19

Community 3: $\left[\frac{4}{20} - \left(\frac{10}{40}\right)^2\right]$ = 0.1375

$$M = \sum_{c=1}^{n_c} \left[\frac{L_c}{L} - \left(\frac{k_c}{2L}\right)^2\right]$$

M= 0.505

MODULARITAS



- **Partisi optimal**, yang memaksimalkan modularitas.
- **Sub-optimal** tetapi positif.
- **Zero/single modularity**: Menempatkan semua node ke *community* yang sama
- **Modularitas Negatif**: Jika setiap node berbeda *community*

ALGORITME DETEKSI KOMUNITAS

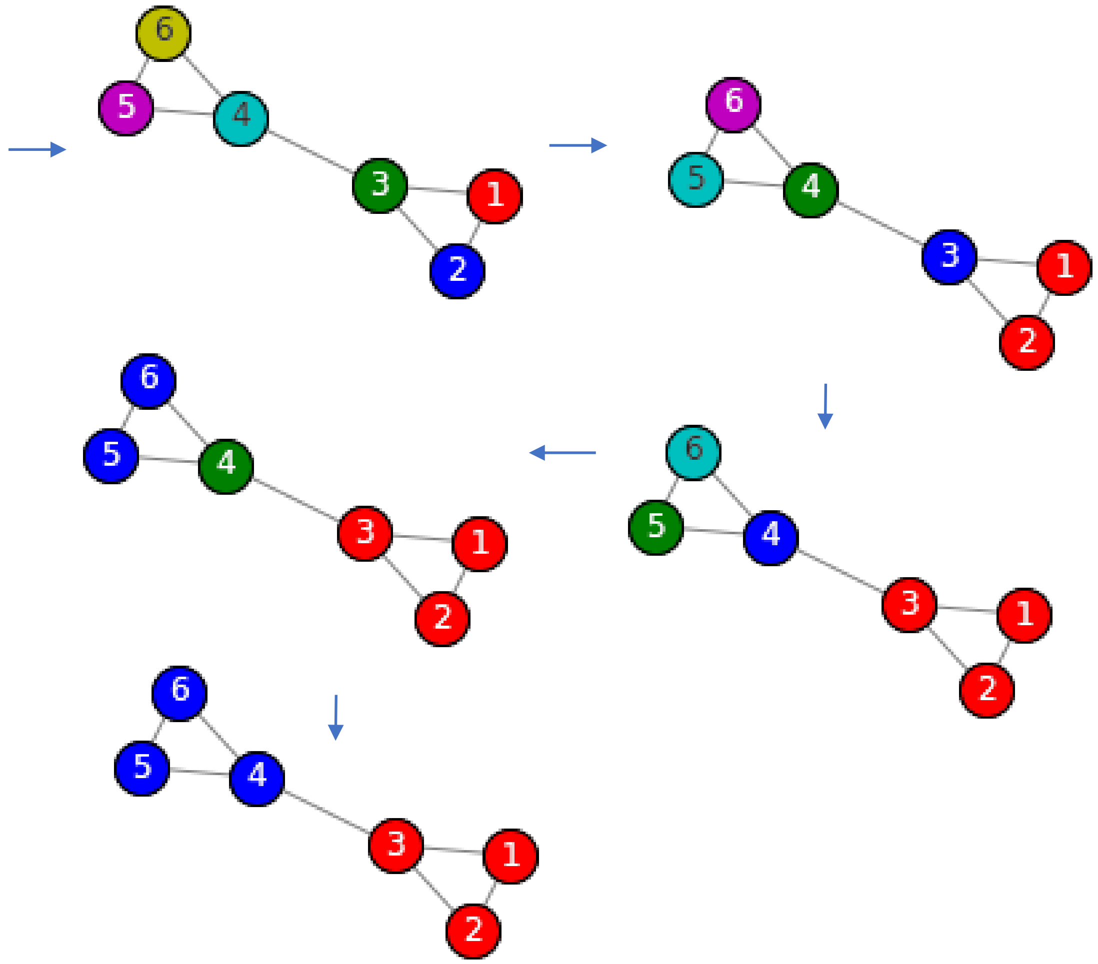
Table 1 Brief summarization of community detection algorithms

Type of community detection algorithms	Author (year)	Approach	Parameters	Type of community/type of network
Modularity algorithms	Newman [1]	Modularity maximization	Modularity	Non-overlapping/Static network
	Clauset et al. [22]	Greedy optimization of Modularity	Edges, vertices, modularity	Non-overlapping/Static network
	Yu and Ding [23]	Normalized modularity	Modularity, spectral clustering	Non-overlapping/static network
	Aggarwal et al. [24]	Optimized modularity using heuristic search	λ—function	Non-overlapping/Static Network
Information theoretic algorithms	Lancichinetti et al. [27]	Local optimization of the fitness function	Local fitness measure	Non-overlapping/static network
	Xie et al. [28]	Vector space model	Cosine function	Non-overlapping/weighted network
	Guan-yu et al. [29]	Mapping vertex into vector	Vector	Non-overlapping/Static Network
	Huang et al. [31]	Local tightness expansion	Tightness function, similarity measure	Non-overlapping/dynamic network
Network algorithms	Newman et al. [37]	CONGA	Edge betweenness, split betweenness	Non-overlapping/static network
	Dang et al. [38]	Vertex similarity	K-nearest neighbor	Non-overlapping/weighted Network
	Picciardi et al. [41]	Edge betweenness	Time series	Non-overlapping/weighted Network
Hierarchical algorithms	Raghwan et al. [42]	Label propagation	Node, edges	Diffusion/dynamic network
	Guang Xu [43]	Latent community discovery	Pareto principle	Overlapping/dynamic network
	Palla et al. [45]	K-Clique		Overlapping/static network

Mittal, R., Bhatia, M.P.S. Classification and Comparative Evaluation of Community Detection Algorithms. *Arch Computat Methods Eng* **28,** 1417–1428 (2021). https://doi.org/10.1007/s11831-020-09421-5

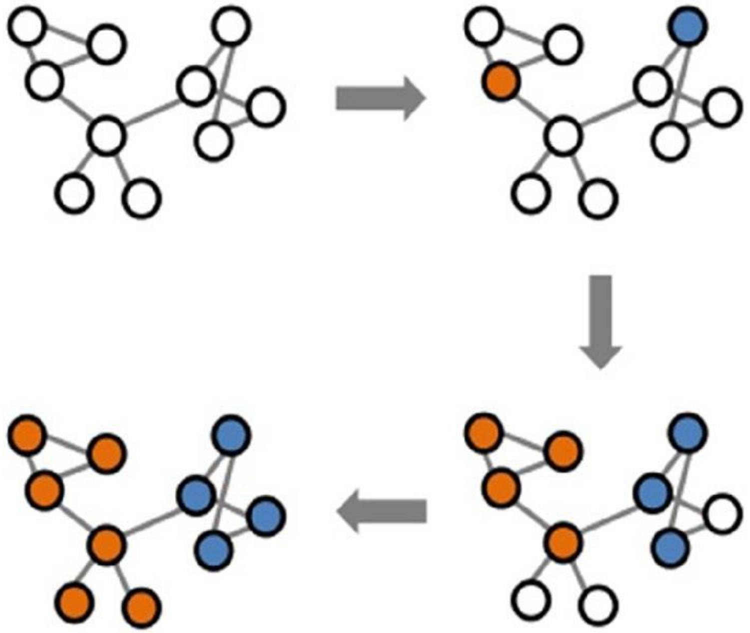
CONTOH ALGORITME DETEKSI KOMUNITAS

Greedy Modularity



Mittal, R., Bhatia, M.P.S. Classification and Comparative Evaluation of Community Detection Algorithms. *Arch Computat Methods Eng* **28,** 1417–1428 (2021). https://doi.org/10.1007/s11831-020-09421-5

Label Propagation



Label propagation algoritme semi-supervised machine learning yang melabeli data yang tak berlabel berdsarkan subset data yang memiliki labelto previously unlabeled data points.

MATERIAL, METODE, DAN HASIL

AKUISISI DATA

- Akuisisi data
  - List penyakit melalui *text mining* naskah di PubMed melalui Pubtator Central (PTC) https://www.ncbi.nlm.nih.gov/research/pubtator/
  - PTC melakukan anotasi otomatis untuk memberikan enam biokonsep: penyakit, gen, spesies, mutasi, kimia, dan garis sel (Wei et al. 2019).
- Praproses data
  - Membersihkan data dan mengidentifikasi komorbiditas; menghapus kata/penyakit: death, mortality, etc.
  - Mencari DOID melalui https://disease-ontology.org/

Table 1 Disease ontology data on pneumonia

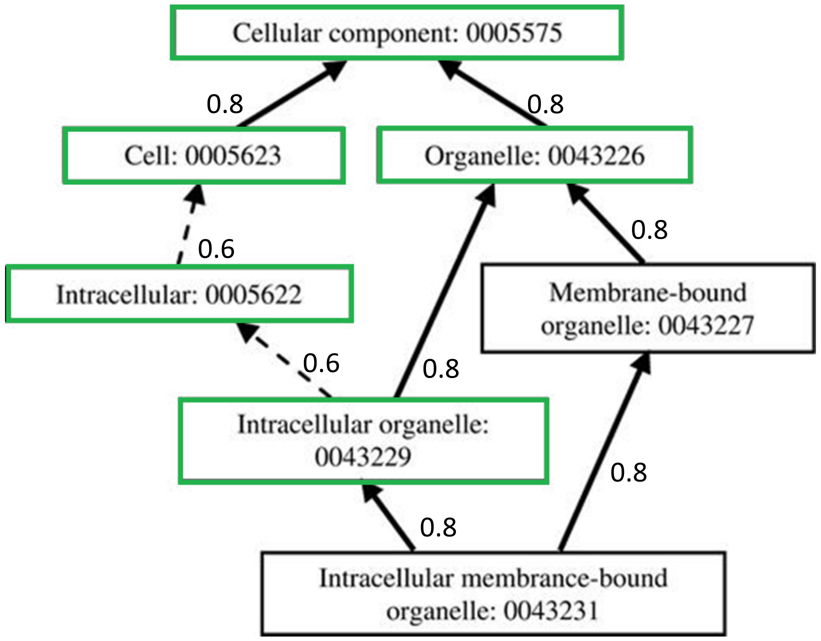
Metadata	Data
ID	DOID:552
Name	Pneumonia
Alternates	DOID:10509 DOID:11742 DOID:5871
Synonym	Acute pneumonia [EXACT]

It has one main DOID and three alternates DOID. All of them refer to same disease

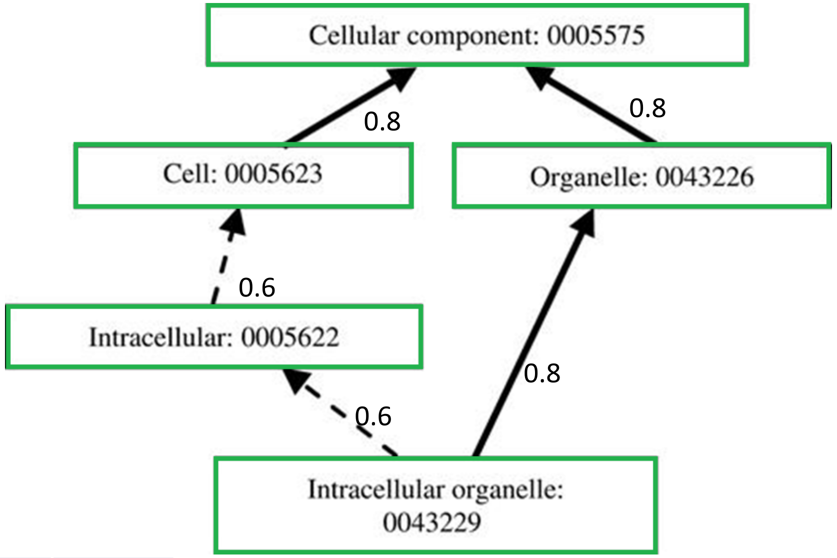
METODE

- Membangun jejaring berdasarkan perhitungan similaritas antar penyakit
  - Perhitungan matriks similaritas menggunakan library DOSE, bahasa R
  - Menggunakan 5 algoritme: Wang, Jiang, Lin, Resnik dan Rel.
  - Elemen matriks berkisar dari 0 hingga 1; semakin mendekati 1 berarti semakin similar;
    - 0: dengan 0 menunjukkan bahwa kedua komorbiditas tidak identik dan 1 menunjukkan bahwa keduanya. Selanjutnya, matriks dianalisis menggunakan ambang batas yang diterapkan.
  - Dengan ambang batas tertentu; dibangun network menggunakan Cytoscape

ILUSTRASI PERHITUNGAN SIMILARITAS



$$\begin{cases} S_A(A) = 1 \\ S_A(t) = \max\{w_e * S_A(t') | t' \in \text{childrenof}(t)\} \text{ if } t \neq A \end{cases}$$



Is A : 0.8

Part of : 0.6

GO terms	SV (A)
0043231	1
0043229	0.8
0043227	0.8
0005622	0.48
0005623	0.288
0043226	0.64
0005575	0.512
SV(A)	4.52
SA(t)	2.72

A ∩ B

GO terms	SV(B)	
0043229	1	
0005622	0.8	
0005623	0.36	= 0.6 * 0.6
0043226	0.8	
0005575	0.64	= max (0.6 * 0.6 * 0.8, 0.8 *0.8)
SV(B)	3.44	
SB(t)	3.44	

$$S_{GO}(A,B) = \frac{\sum_{t \in T_A \cap T_B} (S_A(t) + S_B(t))}{SV(A) + SV(B)}$$

=(2.72 +3.44)/(4.52+3.44)

= **0.7727**

$$SV(A) = \sum_{t \in T_A} S_A(t)$$

JEJARING HASIL PERHITUNGAN SIMILARITAS PENYAKIT

- Network didasarkan perhitungan matriks similarity; 338 nodes
- Ditentukan komunitas pada **giant component**

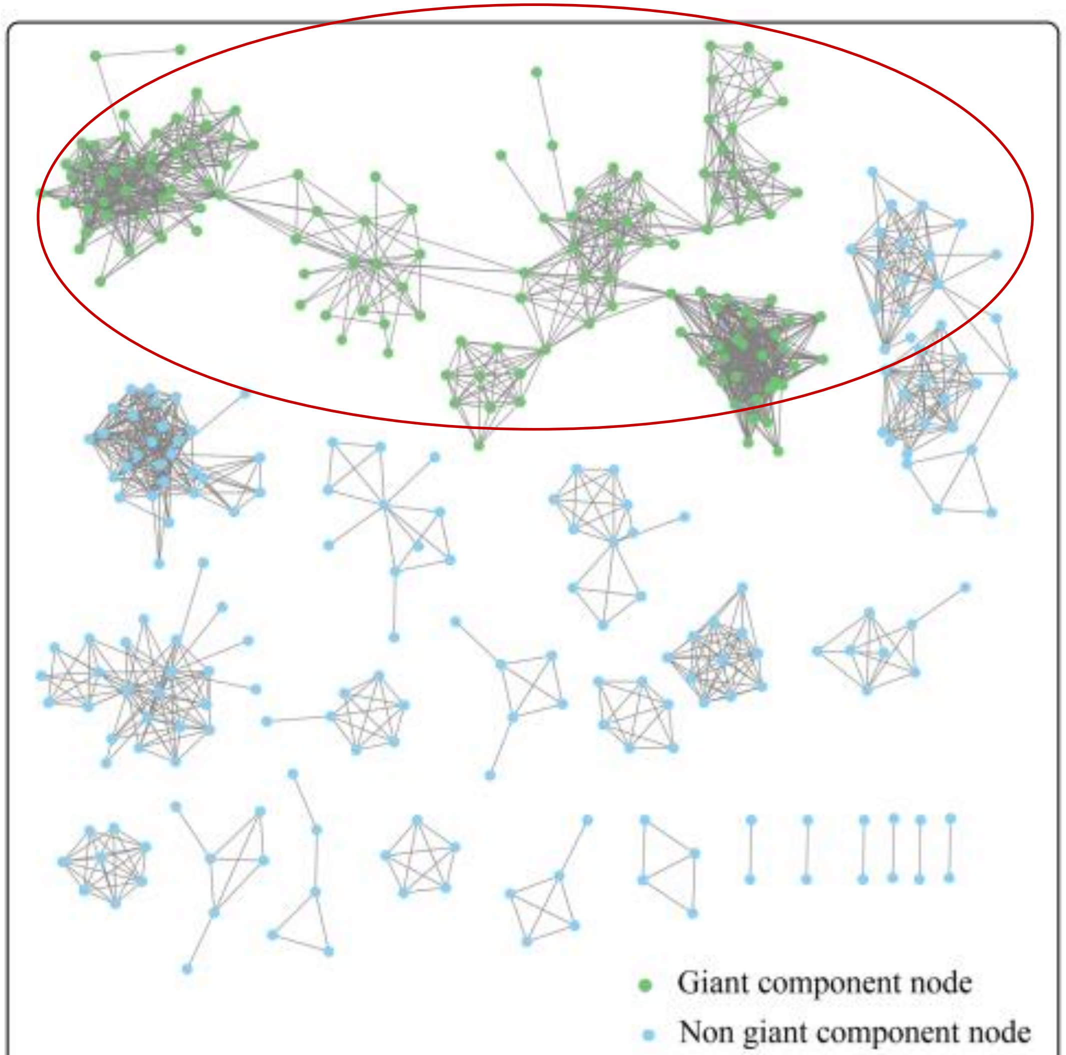


Fig. 3 Network formation using Cytoscape. Each node represents a disease, and the link represents a similar relationship between diseases. There is a giant component, with the number of nodes = 144

METODE

- Menentukan komunitas menggunakan 20 algoritme dan menghitung 5 modularity
- Menentukan penyakit komorbid signifikan pada setiap komunitas, menggunakan betweenness, degree, closeness, dan eigenvector centrality.

$$Q(S)_{NewmanGirvan} = \frac{1}{m} \sum_{c \in S} \left(m_s - \frac{(2m_s + l_s)^2}{4m} \right) \quad (1)$$

$$Q(S)_{ErdosRenyi} = \frac{1}{m} \sum_{c \in S} \left(m_s - \frac{(mn_s(n_s - 1))}{n(n-1)} \right) \quad (2)$$

$$Q(S)_{LinkModularity} = \frac{1}{2m} \sum_{i,j \in V} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \quad (3)$$

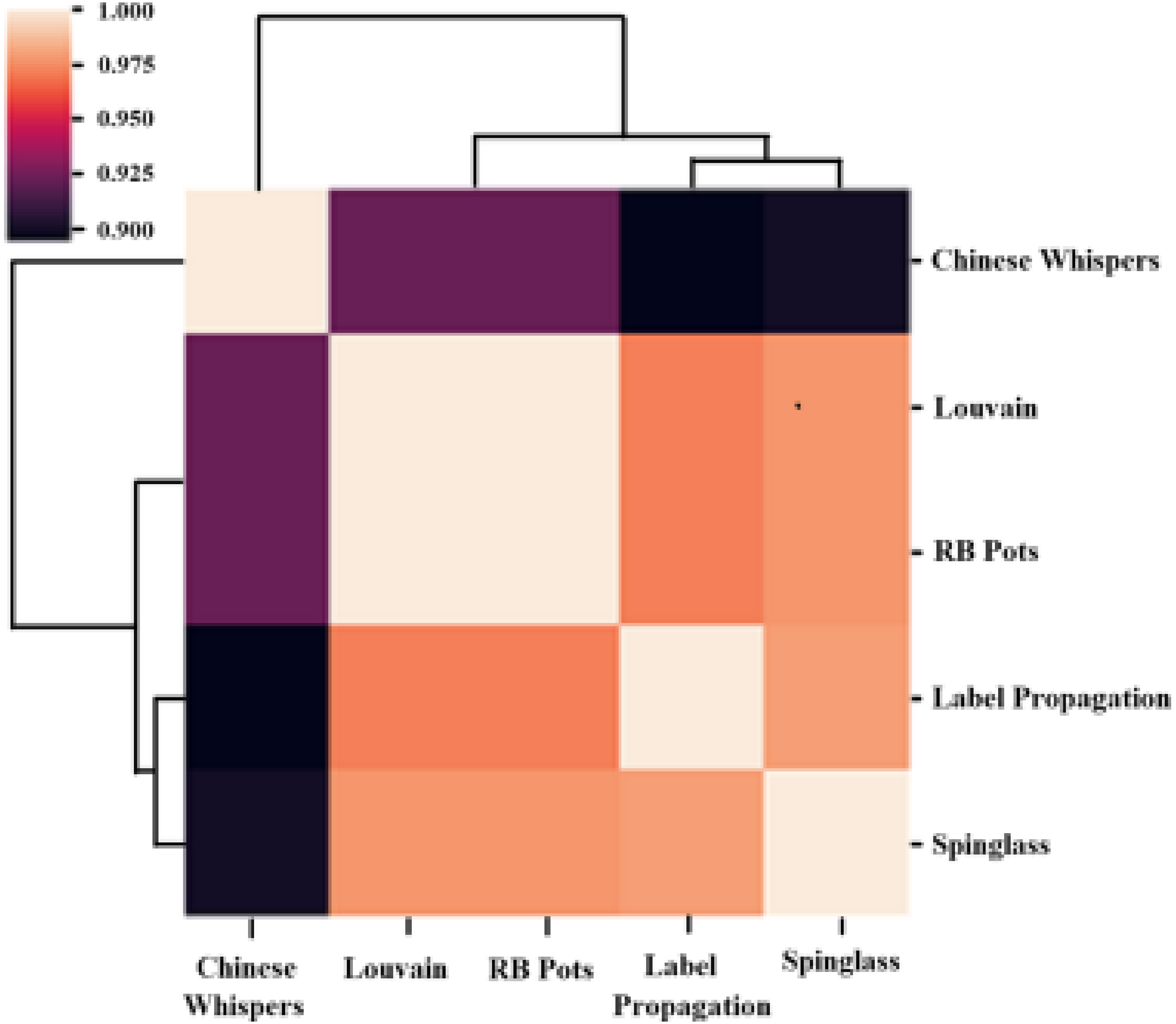
$$Q(S)_{ModularityDensity} = \sum_{c \in S} \frac{1}{n_c} \left(\sum_{i \in C} 2 * \lambda * k_{iC}^{int} - \sum_{i \in C} 2 * \lambda * k_{iC}^{out} \right) \quad (4)$$

$$Z(C)_{ZModularity} = \frac{\sum_{c \in C} \frac{m_c}{m} - \sum_{c \in C} (\frac{D_c}{2m})^2}{\sqrt{\sum_{c \in C} (\frac{D_c}{2m})^2 (1 - \sum_{c \in C} (\frac{D_c}{2m})^2)}} \quad (5)$$

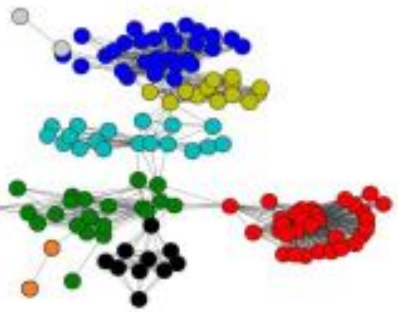
LIMA ALGORITME DETEKSI KOMUNITAS TERBAIK

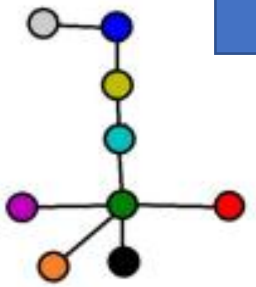
Algoritme	Newman Grvan	Erdos Renyi	link	density	Z
Label Propagation	0.710	0.781	0.134	73.272	1.716
Spinglass	0.710	0.781	0.134	73.272	1.716
Chinese whispers	0.711	0.773	0.135	70.804	1.715
Louvain	0.711	0.773	0.135	70.804	1.715
RB Pots	0.711	0.773	0.135	70.804	1.715

- Komunitas ditentukan menggunakan library cdlib
- Ada lima algoritma modularitas yang digunakan, yaitu Newman Girvan, Erdos Renyi Modularity, Link Modularity, Modularity Density, dan Z Modularity
- Lima algoritme terbaik: Label Propagation, Spinglass, Chinese berbisik, Louvain, dan RB Pots

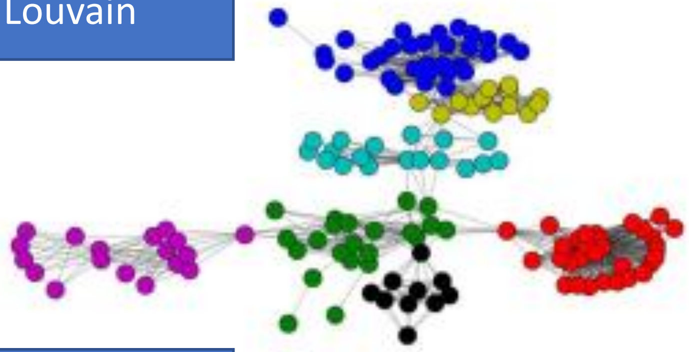

1.000
0.975
0.950
0.925
0.900
Chinese Whispers
Louvain
RB Pots
Label Propagation
Spinglass
Chinese Whispers
Louvain
RB Pots
Label Propagation
Spinglass

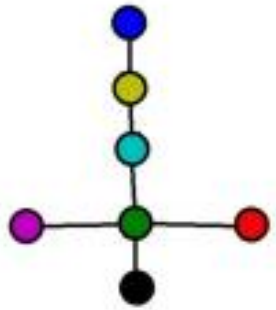
Label Propagation



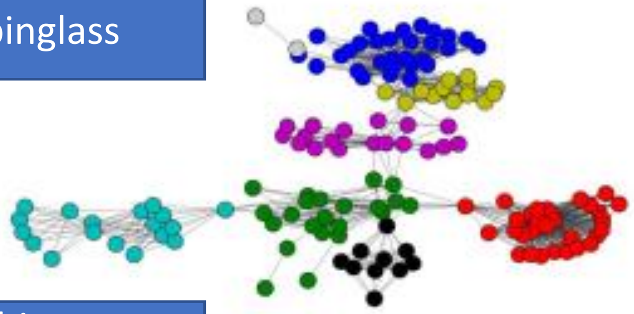


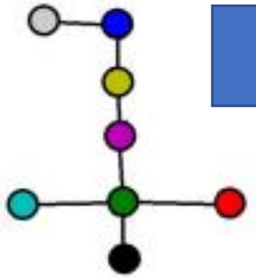
Louvain



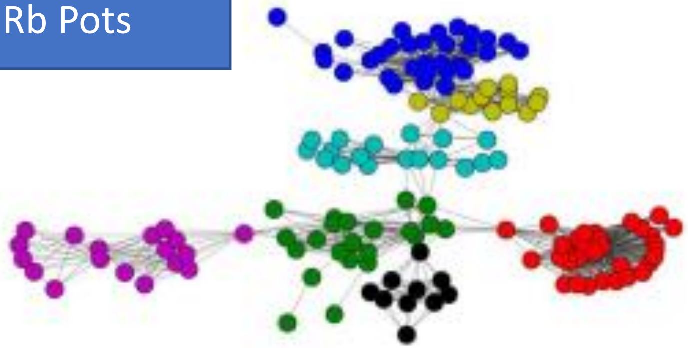


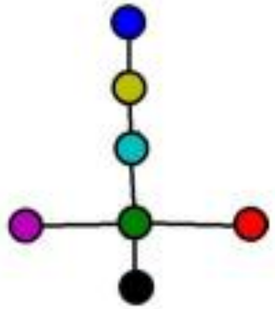
Spinglass



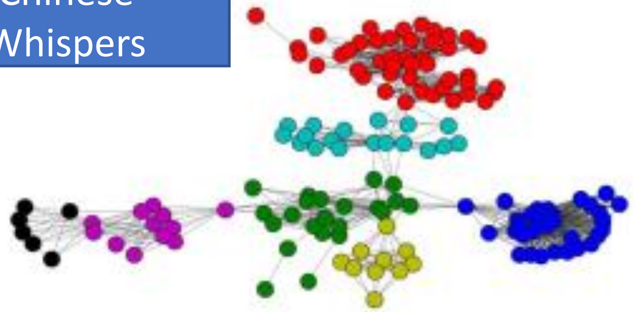


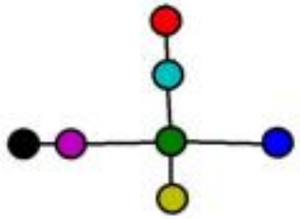
Rb Pots





Chinese Whispers





- Louvain: algoritme deteksi komunitas yang menggunakan pendekatan optimisasi greedy untuk mengoptimasi modularitas
- Chinese whispering: metode hard clustering, randomized, and flat clusetering
- Spinglass mengadopsia statistical mechanics approach, dengan mendefinisikan ukuran baru untuk kekuatan komunitas
- RbPost melakukan optimisasi fungsi kualitas yang didefinisikan

DAFTAR PENYAKIT KOMORBID HASIL DETEKSI KOMUNITAS

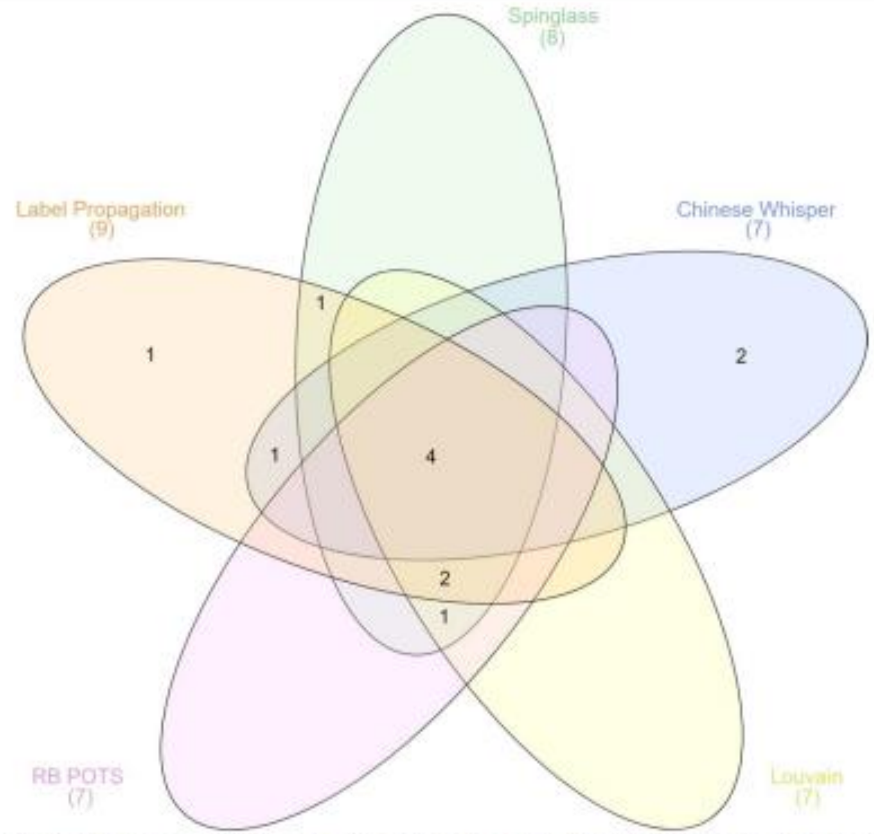


Fig. 7 Venn diagram among the five algorithms. The numbers indicates how many comorbid diseases there are

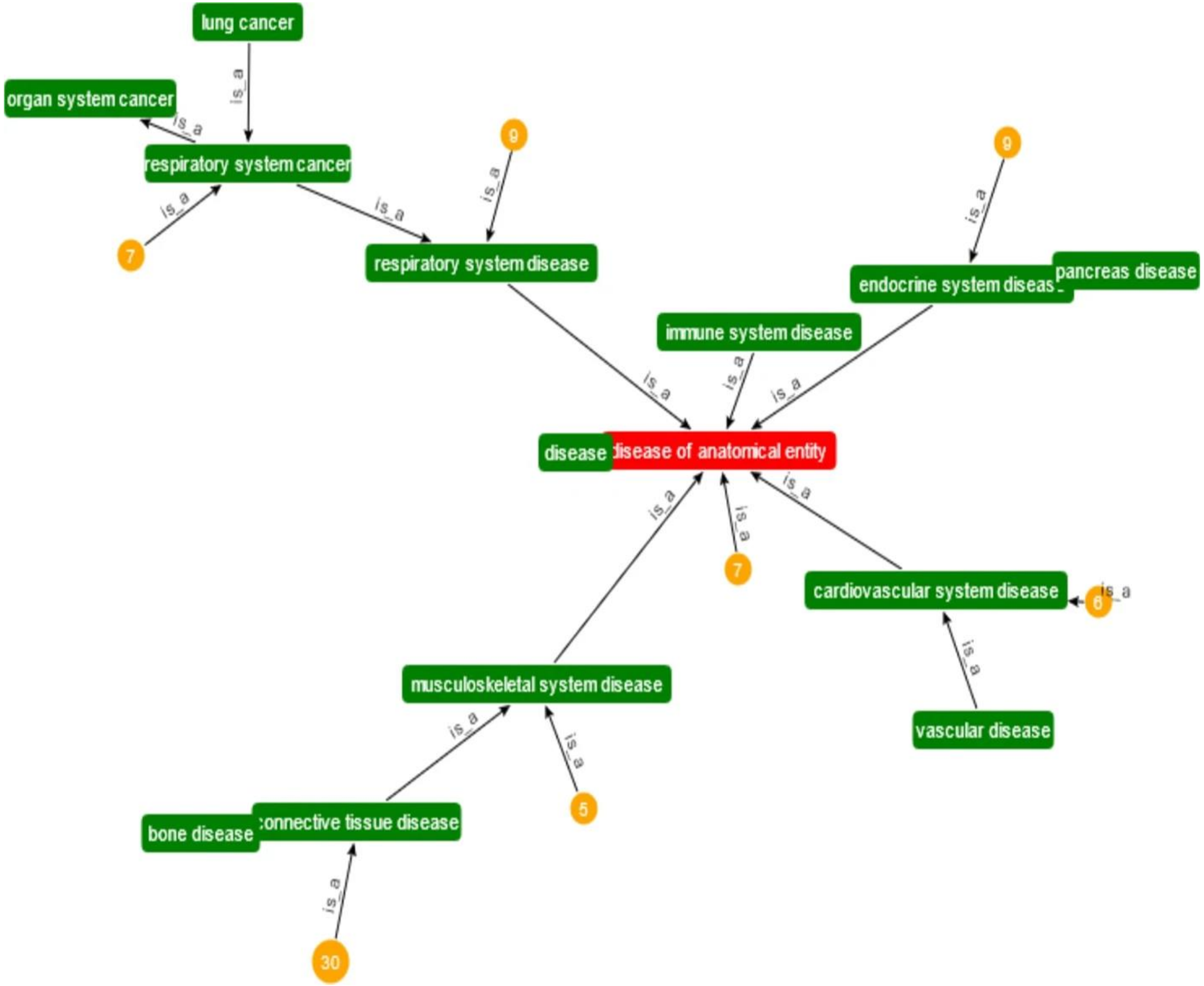
Table 7 Algorithm and significant disease

Algorithm	Disease
Label Propagation and Spinglass and Chinese whisper and Louvain and RB POTS	Vascular disease, immune system disease, bone disease, pancreas disease
Label Propagation and Spinglass and Louvain and RB POTS	Disease of metabolism, atrial heart septal defect
Spinglass and Louvain and RB POTS	Interstitial lung disease
Label Propagation and Spinglass	Familial atrial fibrillation
Label Propagation and Chinese whisper	Respiratory system disease
Chinese whisper	Diabetes mellitus, familial hyperlipidemia
Label Propagation	Persistent generalized lymphadenopathy

There are various lists of the same disease found in different algorithms

Disease Ontology

Struktur Directed Acyclic Graph pada vascular, immune system, bone, pancreas disease, dan kanker paru dapat dinyatakan dalam relasi pada disease ontology.



Kesimpulan

1. Network penyakit yang dikembangkan berdasar kesamaan ontologi penyakit dapat digunakan untuk deteksi komorbiditas dengan algoritme deteksi komunitas
2. Algoritme yang relevan dalam pengelompokan komorbiditas kanker paru adalah: Label propagation, Spinglass, Chinese Whispering, Louvain, dan RB Pots
3. Komorbiditas yang signifikan adalah: vascular disease, immune system disease, bone disease, pancreatic disease

Terima Kasih

Heru Cahya Rustamaji

Yustina Sri Suharini

Angga Aditya Permana

Wisnu Ananta Kusuma

Sri Nurdiati

Irmanida Batubara

Taufik Djatna

IPB University
— Bogor Indonesia —

Department of Computer Science
FMIPA-IPB Kampus Darmaga
Jl Meranti Wing 20 Level V, Bogor, Indonesia
Phone/Fax: +62 25v 8625584
http://cs,ipb,ac,id/