

Disjoint Community Detection pada Network Kegiatan Kemahasiswaan di ISTTS Menggunakan Fast Greedy dan Walktrap

Mikhael Setiawan, *Teknologi Informasi Institut Sains dan Teknologi Terpadu Surabaya,*
Gunawan, *Teknologi Informasi Institut Sains dan Teknologi Terpadu Surabaya,*
F.X.Ferdinandus, *Teknologi Informasi Institut Sains dan Teknologi Terpadu Surabaya.*

Abstrak—Disjoint community detection bertujuan untuk menemukan sebuah komunitas pada network dengan melakukan pemisahan. Pada penelitian ini, disjoint akan dilakukan pada network kegiatan kemahasiswaan di ISTTS. Metode disjoint community detection yang digunakan adalah fast greedy dan walktrap algorithm. Data kegiatan kemahasiswaan berisi mengenai mahasiswa bersama-sama dengan mahasiswa lainnya mengikuti kegiatan kemahasiswaan apa saja. Setelah disjoint berhasil dilakukan, maka akan dihitung nilai closeness centrality dari setiap mahasiswa, dimana pada akhirnya akan dihitung correlation coefficient dengan IPK mahasiswa tersebut untuk mencari hubungan antara centrality mahasiswa dengan IPK mereka. Hasil closeness centrality ini selanjutnya di rata-rata untuk semua hasil algoritma untuk melihat bagaimana korelasi closeness centrality dengan ipk mahasiswa tersebut. Uji coba dilakukan dengan membentuk gml dari kombinasi filter, yang menghasilkan sekitar 2527 gml dengan nilai akhir korelasi adalah 62 - 63% weak positif dengan diikuti 16-18% moderate positif, dan 14-16% tidak berkorelasi sama sekali. Akhirnya dapat disimpulkan bahwa closeness centrality dalam sebuah komunitasnya, hanya berpengaruh secara weak positif dengan ipk mahasiswa tersebut.

Kata Kunci— Closeness Centrality, Correlation Coefficient, Disjoint Community Detection, Network Kegiatan Mahasiswa.

I. PENDAHULUAN

Mahasiswa yang menjalani pendidikan pada sebuah Universitas tidak hanya dituntut untuk berprestasi secara akademis namun juga dituntut memiliki soft skill. Pada universitas ISTTS yang merupakan studi kasus dalam penelitian ini, pengembangan soft skill mahasiswa dibuat dalam bentuk kegiatan dan organisasi kemahasiswaan seperti kegiatan kerohanian, pengembangan minat dan bakat, pengembangan penalaran mahasiswa, dan keorganisasi seperti BEM maupun HIMA.

Ilmu pada social network analysis telah berkembang

sangat pesat, salah satunya pada cabang ilmu community detection. Disjoint community detection adalah salah satu cabang ilmu community detection yang berusaha mencari komunitas-komunitas kecil ini dengan cara dipisahkan. Pada penelitian ini akan digunakan dua jenis disjoint community detection yaitu fast greedy dan walktrap.

Kegiatan bersama yang dilakukan mahasiswa-mahasiswa ini dapat dimodelkan menjadi sebuah network berkegiatan bersama-sama, sebagai contoh bila seorang mahasiswa, anggaplah bernama A sering ditemukan berkegiatan bersama dengan mahasiswa bernama B, maka bisa dianggap kedua mahasiswa ini memiliki hubungan berkegiatan bersama, bahkan lebih jauh kemungkinan besar berteman satu sama lain. Pada penelitian ini, network akan dibangun dari seberapa lama masing-masing mahasiswa berkegiatan bersama dengan mahasiswa lainnya.

Lebih jauh lagi, tidak hanya menemukan komunitas yang lebih kecil, penelitian ini akan berusaha untuk menemukan mahasiswa mana saja yang menjadi central dari komunitas yang lebih spesifik tersebut. Pencarian siapa mahasiswa yang menjadi central dari komunitas tersebut dilakukan dengan melakukan perhitungan nilai closeness centrality, semakin besar nilai closeness ini maka mahasiswa dianggap semakin aktif dan merupakan central dari komunitas.

Hasil nilai closeness centrality mahasiswa dalam komunitas kecil tersebut, nantinya akan coba ditemukan korelasinya dengan indeks prestasi kumulatif atau IPK mahasiswa. Perhitungan korelasi ini akan dilakukan dengan metode pearson correlation coefficient, dengan melihat hubungan antara nilai closeness centrality dengan IPK mahasiswa tersebut.

Beberapa penelitian sejenis yang berusaha untuk menemukan hubungan antara kegiatan kemahasiswaan atau kegiatan ekstrakurikuler diantaranya adalah milik Robert Freeman berjudul the relationship between extracurricular activities and academic achievement[1], dan penelitian milik Gail M. Ritchie dengan judul the impact of academic co-curricular activity participation on academic achievement : a study of catholic high school students[2]. Pada akhirnya penelitian ini bertujuan untuk menjawab beberapa pertanyaan seperti bagaimana hubungan antara kegiatan kemahasiswaan dengan IPK mahasiswa, dan lebih jauh bagaimana hubungan kegiatan kemahasiswaan ini dengan lama study yang ditempuh oleh mahasiswa.

Mikhael Setiawan, Teknologi Informasi Institut Sains dan Teknologi Terpadu Surabaya, Surabaya, Jawa Timur, Indonesia (e-mail : mikhael@stts.edu)

Gunawan, Teknologi Informasi Institut Sains dan Teknologi Terpadu Surabaya, Surabaya, Jawa Timur, Indonesia (e-mail : gunawan@stts.edu)

F.X.Ferdinandus, Teknologi Informasi Institut Sains dan Teknologi Terpadu Surabaya, Surabaya, Jawa Timur, Indonesia (e-mail : ferdi@stts.edu)

II. TINJAUAN PUSTAKA

Pada bagian ini akan dijelaskan beberapa riset yang menjadi landasan teori dari penelitian ini. Fokus utama dari penelitian ini adalah disjoint community detection yang dapat digunakan untuk menemukan komunitas yang lebih kecil dengan melakukan pemisahan[3]. Lebih lanjut, bagian ini akan membahas mengenai algoritma fast greedy dan walktrap sebagai algoritma disjoint community detection, closeness centrality, dan perhitungan korelasi dengan pearson correlation coefficient.

A. Fast Greedy Algorithm

Fast greedy algorithm merupakan pengembangan dari greedy optimization milik Newman. Greedy optimization milik Newman merupakan agglomerative hierarchical clustering, yang termasuk dalam modularity based algorithm dengan tujuan untuk mengoptimalkan nilai modularity Q [4][5]. Nilai modularity Q dapat dihitung dengan menggunakan rumus berikut.

$$Q = \sum_{c=1}^{n_c} \left[\frac{l_c}{m} - \left(\frac{d_c}{2m} \right)^2 \right] \quad (1)$$

Dimana:

- n_c = jumlah cluster
- l_c = total jumlah edges yang menggabungkan vertices dari cluster c
- d_c = sum dari degree pada vertices cluster c
- $\frac{l_c}{m}$ = fraction dari edges pada graph dalam cluster c
- $\left(\frac{d_c}{2m} \right)^2$ = fraction dari edges yang diharapkan pada null model

Nilai modularity ini yang nantinya digunakan untuk menentukan apakah subgraph yang terbentuk dari disjoint community detection merupakan komunitas yang baik atau tidak[6]. Proses yang dilakukan oleh greedy optimization milik Newman dimulai dengan menganggap setiap vertex merupakan anggota tunggal dalam sebuah komunitas. Lalu kedua komunitas ini digabung lalu dihitung nilai modularity Q nya dengan mencari peningkatan nilai modularity Q atau setidaknya bila nilai Q turun, dicari yang penurunannya paling rendah. Proses penggabungan dan penghitungan nilai modularity Q ini dilakukan terus menerus hingga semua vertex menjadi satu buah komunitas, lalu dilakukan proses pemotongan. Pemotongan dilakukan pada iterasi dengan nilai Q paling maksimal.

Beberapa improvement dari segi jenis struktur data, yang ditawarkan pada fast greedy algorithm oleh Clauset adalah menyimpan matrix berisi ΔQ_{ij} untuk pasangan vertex ij , dimana setiap baris dari matrix disimpan sebagai balanced binary tree. Max-heap H digunakan untuk menyimpan element terbesar dari setiap row matrix ΔQ_{ij} beserta dengan label ij sesuai dengan pasangan pada community-nya. Sebuah vector array yang berisi elements a_i , dimana $a_i = \frac{k_i}{2m}$.

Dari jenis struktur data yang ditawarkan itu, maka alur kerja dari fast greedy, dimulai dari menghitung nilai awal ΔQ_{ij} dan a_i , lalu mengisi max-heap H dengan element terbesar dari setiap baris pada matrix ΔQ . Selanjutnya dipilih ΔQ_{ij} terbesar dari max-heap H , lalu gabungkan kedua communities tersebut, lalu update nilai ΔQ , heap H ,

dan a_i , serta naikkan nilai Q akhir dengan ΔQ_{ij} . Langkah ini dilakukan terus hingga muncul 1 komunitas terakhir. Berikut adalah persamaan untuk nilai ΔQ_{ij} dan a_i .

$$\Delta Q_{ij} = \begin{cases} \frac{1}{2m} - \frac{k_i k_j}{(2m)^2} & \text{bila vertex } i, j \text{ terhubung} \\ 0 & \text{bila vertex } i, j \text{ tidak terhubung} \end{cases} \quad (2)$$

$$a_i = k_i / 2m \quad (3)$$

B. Walktrap Algorithm

Algoritma Walktrap yang diperkenalkan oleh Pascal Pons dan Matthieu Latapy ini merupakan pengembangan dari algoritma random walk pada graph, dengan dasar pemikiran bahwa random walks pada graph akan cenderung untuk terjebak pada bagian pada graph yang terkoneksi secara erat (communities). Pascal dkk menggunakan random walks ini untuk menentukan structural similarity antara vertices dan communities, disebut juga sebagai distance[7]. Nantinya, distance ini akan digunakan pada hierarchical clustering algorithm dan membentuk dendrogram pada akhirnya.

Dalam penentuan komunitasnya, walktrap memiliki sebuah nilai inputan yaitu length (t), selain itu walktrap algorithm juga menggunakan jarak r antara vertex nya sebagai perhitungan. Nilai jarak r ini akan semakin besar bila kedua vertices berada di dua community yang berbeda, dan sebaliknya. Dalam penentuan distance r , Pascal dkk merumuskan nilai probabilitas P_{ik}^t .

Menurut Pascal dkk, beberapa catatan yang perlu diingat mengenai probabilitas P_{ij}^t ini adalah jika vertex i dan j berada dalam 1 komunitas maka nilai P_{ij}^t akan tinggi, probabilitas ini dipengaruhi juga oleh degree $d(j)$ karena semakin tinggi nilai degree maka semakin tinggi pula kemungkinan random walker jalan ke vertex tersebut, dan terakhir dua vertex dalam sebuah komunitas memiliki kemungkinan melewati vertex lain dengan cara yang sama sehingga kemungkinan $\forall k, P_{ik}^t \approx P_{jk}^t$. Dari pemahaman tersebut maka dirumuskan nilai distance / jarak r antara vertices adalah sebagai berikut.

$$r_{ij} = \sqrt{\sum_{k=1}^n \frac{(P_{ik}^t - P_{jk}^t)^2}{d(k)}} \quad (4)$$

Setelah distance antara vertices didapatkan, maka dapat digunakan agglomerative hierarchical clustering untuk melakukan disjoint community detection. Walktrap dimulai dengan menganggap semua vertex adalah anggota tunggal 1 buah komunitas, setelah itu lakukan perhitungan jarak r untuk semua vertices yang saling bertetangga. Setelah perhitungan jarak r , lakukan penggabungan dua buah komunitas lalu update semua nilai jarak r setelah penggabungan communities tersebut. Langkah ini dilakukan hingga ditemukan 1 buah komunitas terakhir.

Untuk pemotongan dendrogram komunitas yang terbentuk, Pascal mencetuskan sebuah nilai rasio η_k sebagai alternatif nilai modularity Q yang telah umum dipakai. Namun pada akhirnya nilai modularity Q tetap menjadi pilihan karena selain sudah umum dipakai, Pascal menjelaskan bahwa nilai rasio η_k hampir sama baiknya dengan nilai modularity Q dalam menentukan best cut dari dendrogram yang dihasilkan.

C. Closeness Centrality

Closeness centrality adalah salah satu perhitungan centrality pada social network analysis. Closeness centrality

digunakan untuk menghitung sum jarak dari satu node ke node lainnya[8],[9]. Node/ vertex dengan nilai closeness centrality yang tinggi berarti memiliki jarak paling pendek dengan semua nodes lainnya.

Nilai closeness centrality untuk sebuah node merupakan $1/\text{sum}(\text{jarak node tersebut ke semua node lainnya})$, tidak jarang nilai closeness centrality ini di normalisasi menjadi $(\text{jumlah nodes} - 1)/\text{sum}(\text{jarak node tersebut ke semua node lainnya})$.

$$C(N_i) = 1/[\sum_{j=a}^n d(N_i, N_j)] \quad (i \neq j) \tag{5}$$

$$\text{Norm. } C(N_i) = \frac{n-1}{[\sum_{j=a}^n d(N_i, N_j)]} \quad (i \neq j) \tag{6}$$

Untuk kasus weighted graph, perhitungan closeness centrality tidak mengalami banyak perubahan, kecuali untuk kasus dimana bobot dari edge malah menunjukkan kedekatan/keeratan kedua node tersebut. Pada kasus ini, maka untuk perhitungan closeness centrality nya, semua bobot pada edge harus di-inversed (1/bobot).

D. Pearson Centrality dan Fisher z-transformation

Pearson correlation merupakan perhitungan untuk menentukan apakah 2 variable saling berkorelasi. Beberapa skenario yang dihasilkan dari pearson correlation diantaranya adalah (1) seiring dengan naiknya nilai variable 1, nilai variable 2 juga ikut naik, (2) seiring dengan naiknya nilai variable 1, nilai variable 2 malah turun, (3) kenaikan variable 1 tidak seiring/berlawanan, sehingga sering disebut kondisi tidak ada korelasi[10].

Correlation Coefficient akan mengembalikan nilai antara -1 sampai 1, dimana nilai -1 menandakan strong negatif dan nilai 1 menandakan strong positif. Detail pemetaan intepretasi terhadap nilai correlation coefficient dapat dilihat pada table berikut.

TABEL 1
INTERPRETASI NILAI CORRECAION COEFFICIENT

Nilai Coefficient Correlation	Kekuatan	Positif/Negatif
-1 s/d -0.90	Very strong	Negatif
-0.89 s/d -0.70	Strong	Negatif
-0.69 s/d -0.40	Moderate	Negatif
-0.39 s/d -0.10	Weak	Negatif
-0.09 s/d 0.09	Tidak ada korelasi	
0.10 s/d 0.39	Weak	Positif
0.40 s/d 0.69	Moderate	Positif
0.70 s/d 0.89	Strong	Positif
0.90 s/d 1	Very Strong	Positif

Dari hasil correlation coefficient yang didapat, terkadang ingin diambil nilai rata-rata correlation coefficientnya. Hal ini tidak dapat dilakukan dengan cara rata-rata biasa (mean) karena nilai r akan bias menuju nilai 0, sehingga untuk menyelesaikan masalah bias ini, maka digunakan fisher z-transformation. Pertama-tama, masing-masing nilai r akan diubah menjadi nilai fisher's z :

$$z' = 0.5 \ln \frac{1+r}{1-r} = \text{artanh}(r) \tag{7}$$

Selanjutnya nilai z' ini dapat di rata-rata dan hasil dari rata-rata tersebut akan di ubah kembali menjadi nilai r :

$$r_{z'} = \frac{e^{2z'} - 1}{e^{2z'} + 1} = \tanh(z) \tag{8}$$

Nilai $r_{z'}$, inilah yang akan menjadi nilai rata-rata dari correlation coefficient-nya[11].

E. NetworkX dan IGraph

Pada penelitian ini, akan ada dua tools yang nantinya akan digunakan yaitu NetworkX dan IGraph. NetworkX adalah sebuah package untuk bahasa pemrograman Python yang dapat digunakan untuk melakukan eksplorasi dan analisa terhadap network dan disediakan pula network algoritma [12]. Pada penelitian ini networkx akan digunakan untuk membentuk graph modelling language (gml) dari sumber data kegiatan kemahasiswaan. Beberapa function yang akan digunakan dari NetworkX diantaranya adalah write_gml dan write_pajak.

Igraph adalah koleksi dari network analysis tools yang mengutamakan efisiensi, portability, kemudahan dalam penggunaannya, dan tersedia untuk bahasa pemrograman R, Python, Mathematica, dan C/C++[13]. Pada penelitian ini, penggunaan igraph dilakukan karena NetworkX sampai pada saat tulisan ini ditulis, tidak memiliki function bantuan untuk algoritma fast greedy dan walktrap, sedangkan igraph memiliki function ini. Beberapa function yang akan digunakan dari igraph diantaranya adalah community_fastgreedy dan community_walktrap dengan parameter yang dapat ditentukan.

F. Vis.js dan Pajak

Pada penelitian ini, digunakan vis.js dan pajak sebagai alat bantu visualisasi pada website. Pajak dipilih sebagai alat bantu visualisasi komunitas yang terbentuk karena performanya yang baik [14]. Untuk pajak, penelitian ini akan menyediakan hasil network dalam bentuk file .net dan hasil disjoint community detection dalam bentuk file .clu

Vis.js adalah salah satu library javascript untuk memvisualisasikan network beserta dengan hasil komunitas yang terbentuk dengan bantuan warna, sehingga dapat ditampilkan langsung pada website [15].

G. Relevant Literature milik Robert Freeman

Salah satu penelitian yang menjadi rujukan adalah penelitian milik Robert Freeman[1]. Pada penelitian ini, independent variable adalah partisipasi siswa pada kegiatan ekstrakurikuler. Dependant variable adalah skor ACT dan prestasi akademik (GPA). Frequency distribution digunakan untuk memvisualisasikan relasi ACT, GPA, dan kegiatan ekstrakurikuler. Metode yang digunakan adalah coefficient correlation antara variable yang telah disebutkan sebelumnya.

Beberapa research question yang ditanyakan pada penelitian Robert diantaranya adalah seberapa signifikan pengaruh dari jumlah dan partisipasi seorang siswa dalam ekstrakurikuler dengan ACT dan GPA yang didapat. Beberapa variable lain yang diperhitungkan adalah pengaruh jenis kelamin, free, reduced lunch status, dan jumlah ekstrakurikuler dengan GPA, namun penelitian ini tidak memperhitungkan variable-variable ini.

Hasil yang didapatkan adalah ketika siswa mengikuti kegiatan ekstrakurikuler, terjadi peningkatan pada GPA yang didapatkan, tidak peduli jenis kelamin dari siswa tersebut. Skor coefficient correlation yang didapatkan adalah 0.33 yang mengindikasi bahwa ada pengaruh weak positif antara GPA dan kegiatan ekstrakurikuler.

H. Relevant Literature milik Gail M. Ritchie

Salah satu penelitian yang menjadi rujukan adalah penelitian milik Gail M. Ritchie[2]. Pada penelitian ini, variable yang diperhitungkan adalah academic co-curricular activity participation (ACCAP) dan GPA. Penelitian ini memiliki research question yang hamper sama dengan milik Robert, yaitu seberapa jauh keterlibatan siswa dalam kegiatan extracurricular dengan academic performance yang didapatkan. Penelitian ini juga melibatkan intensitas keterlibatan dalam ekstrakurikuler, gender, dan etnis, namun kembali lagi, gender dan etnis tidak menjadi variable dalam penelitian ini. Metode penelitian yang dilakukan adalah mencari coefficients dari regression models.

Hasil yang didapatkan adalah semakin lama dan semakin sering seorang siswa mengikuti academic co-curricular activities, maka terjadi peningkatan GPA sebanyak 0.13 point. Gail M. Ritchie juga menjelaskan bahwa pada penelitian sebelumnya yang dilakukan Reeves juga menyatakan bahwa semakin meningkatnya ACCAP, maka performa akademik juga meningkat.

III. METODE DAN ALUR PENELITIAN

Pada bagian ini akan dijelaskan metode penelitian yang digunakan dalam perhitungan korelasi antara closeness centrality mahasiswa hasil dari disjoint community detection. Penjelasan akan meliputi sumber data, representasi network, alur sistem, dan visualisasi hasil.

A. Sumber Data

Sumber data berasal dari kegiatan kemahasiswaan yang terjadi di ISTTS. Data dibagi menjadi 2 bagian yaitu data kegiatan kemahasiswaan dan data organisasi kemahasiswaan. Kegiatan maupun organisasi ini dibagi-bagi menjadi empat kategori yaitu minat bakat, penalaran, organisasi dan kerohanian.

Data kegiatan berkisar 1282 data, yang berisi detail kegiatan seperti kode, nama, tanggal kegiatan dimulai dan berakhir, kategori kegiatan dan organisasi penyelenggara. Data struktur kegiatan yang berisi mahasiswa siapa saja yang mengikuti kegiatan tersebut, berkisar 42.429 data.

Data organisasi berkisar 673 data, yang berisi detail organisasi seperti kode, kategori, nama, tanggal mulai dan berakhirnya organisasi. Organisasi seperti BEM atau HIMA memiliki durasi kerja sekitar satu tahun, sedangkan organisasi seperti unit kegiatan kerohanian (UKK) dan unit kegiatan mahasiswa (UKM) memiliki durasi kerja sekitar satu semester. Data struktur organisasi yang berisi mahasiswa siapa saja yang mengikuti organisasi tersebut, berkisar 15.708 data.

Beberapa data dari sumber data tidak lengkap, misalnya seperti tanggal berakhirnya kegiatan atau keorganisasian tidak ada, sehingga dilakukan beberapa pembenaran data seperti memberikan nilai default satu semester atau satu tahun untuk organisasi-organisasi seperti BEM, HIMA, UKM, ataupun UKK.

B. Representasi Network

Berdasarkan sumber data yang ada, selanjutnya dilakukan proses untuk mengubah data menjadi network, dalam format graph modelling language (gml). Pada penelitian ini, data mahasiswa merepresentasikan node / vertex, sedangkan

untuk edge yang terbentuk dari antara node dilihat dari apakah kedua orang mahasiswa pernah berkegiatan bersama dalam sebuah kegiatan atau tidak.

Hasil dari step ini adalah sebuah gml dari weighted undirected network. Untuk pembentukan gml-nya, digunakan bantuan library NetworkX. Sebagai contoh, bila mahasiswa A dan mahasiswa B pernah berkegiatan bersama dalam kegiatan X dan organisasi Y, maka akan dibuatkan dua buah node yaitu A dan B, lalu kedua node tersebut akan dihubungkan dengan sebuah edge. Berdasarkan contoh, durasi lama kegiatan dan organisasi tersebut berlangsung, dalam satuan hari, akan dijumlah dan dijadikan nilai bobot/weight dari edge penghubung node A dan B. Semakin lama durasi kegiatan dan semakin sering kedua mahasiswa terlibat kegiatan bersama, maka semakin besar nilai weight dari edge. Mengingat nilai bobot ini digunakan untuk ukuran kekuatan hubungan antara node, maka nanti pada saat perhitungan closeness centrality, bobot ini akan di inverted.

```
graph [
  node [
    id 0
    label "217140071"
    mhs_nama "ERIC SAN FOKALIE"
    jur_kode "14"
    jur_nama "S1-DESAIN PRODUK"
    mhs_ipk 3.32
    mhs_poin 1270
    angkatan "2017"
  ]
  node [
    id 1
    label "217140072"
    mhs_nama "GERARD RUSTADY"
    jur_kode "14"
    jur_nama "S1-DESAIN PRODUK"
    mhs_ipk 3.47
    mhs_poin 2957
    angkatan "2017"
  ]
  edge [
    source 0
    target 1
    weight 3
  ]
]
```

Gambar. 1. Contoh gml yang dihasilkan dengan memuat node dan edge

Pada Gambar. 1, dapat dilihat gml akan menyimpan data mahasiswa seperti nama, jurusan, angkatan, dan ipk mahasiswa tersebut dalam daftar node, sedangkan pada daftar edgenya akan berisi node mana saja yang saling berhubungan dengan diberikan informasi weight.

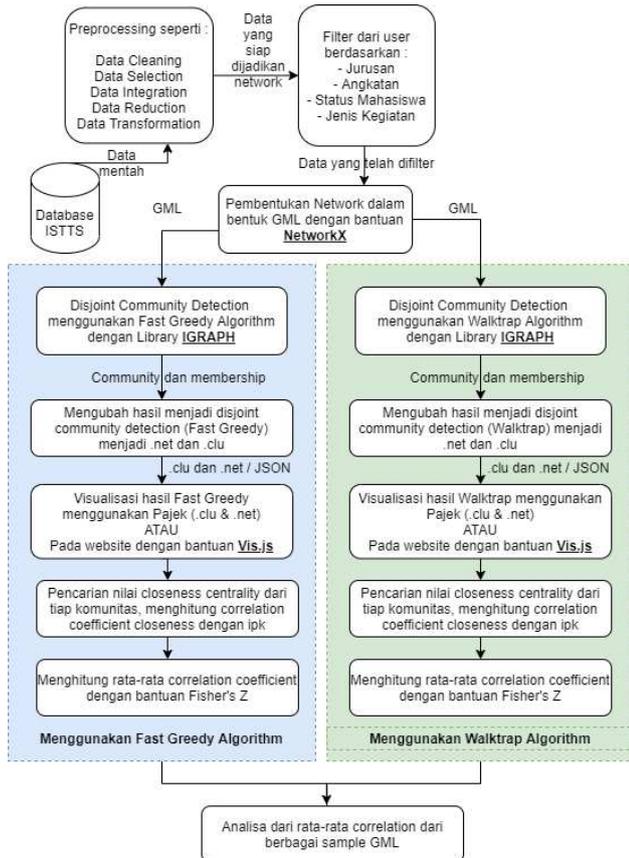
C. Alur Penelitian

Alur penelitian dalam penelitian ini dapat dilihat pada Gambar. 2. Alur akan dijelaskan mulai dari data mentah, preprocessing, pembentukan gml, disjoint community detection, pencarian nilai closeness, pencarian nilai rata-rata closeness, hingga pembuktian hipotesa yaitu ada tidaknya korelasi antara semakin berpengaruhnya seorang mahasiswa dalam komunitas dengan IPK yang dimilikinya.

Alur penelitian ini sedikit berbeda dengan penelitian yang dilakukan oleh Robert dan Gail M. Ritchie. Penelitian Robert hanya melakukan perhitungan correlation, dan penelitian Gail M. Ritchie melakukan regression model. Kedua penelitian rujukan tidak melakukan proses disjoint

community detection, dan hanya melakukan perhitungan korelasi saja.

Pada Gambar. 2, dapat dilihat alur dimulai dengan mendapatkan data dari sumber data yaitu kegiatan kemahasiswaan ISTTS, dilanjutkan dengan preprocessing seperti membersihkan data yang terlalu tua dan tidak valid misalnya data kegiatan tahun 1990-an. Setelah data sudah cukup baik dan dapat digunakan, preprocessing dilanjutkan dengan mengambil hanya kolom-kolom yang perlu saja seperti kategori, tanggal awal, tanggal selesai kegiatan/ organisasi, dan data mahasiswa seperti nrp, nama, jurusan, dan angkatan.



Gambar. 2. Block Diagram Alur Kerja Penelitian

Setelah data siap, lalu dilakukan pembentukan network GML seperti yang telah dijelaskan pada subbab sebelumnya. Pembentukan gml ini akan dipengaruhi oleh filter yang dimasukkan oleh pengguna. Filter pencarian yang disediakan adalah jurusan, angkatan, status mahasiswa, dan jenis kegiatan. Jurusan adalah jurusan mahasiswa tersebut seperti informatika, industri, dan sebagainya. Angkatan adalah angkatan mahasiswa masuk. Status mahasiswa disini adalah aktif tidak nya mahasiswa tersebut, dimana nilai 0 menandakan mahasiswa tersebut tidak aktif/ dropout, nilai 1 menandakan mahasiswa tersebut aktif, nilai 9 menandakan mahasiswa tersebut telah lulus/ menjadi alumni. Jenis kegiatan/ organisasi dibagi menjadi 4 kategori yaitu minat bakat, organisasi, penalaran, dan kerohanian.

GML yang terbentuk selanjutnya akan melalui proses disjoint community detection, seperti yang dapat dilihat pada Gambar. 2, untuk disjoint dapat dipilih ingin menggunakan fast greedy atau walktrap. Dari hasil disjoint

community detection baik menggunakan fast greedy maupun walktrap, akan didapat komunitas yang dianggap baik. Dari setiap komunitas tersebut akan dibentuk sebuah subgraph lalu dihitung nilai closeness centrality-nya untuk masing-masing node mahasiswa.

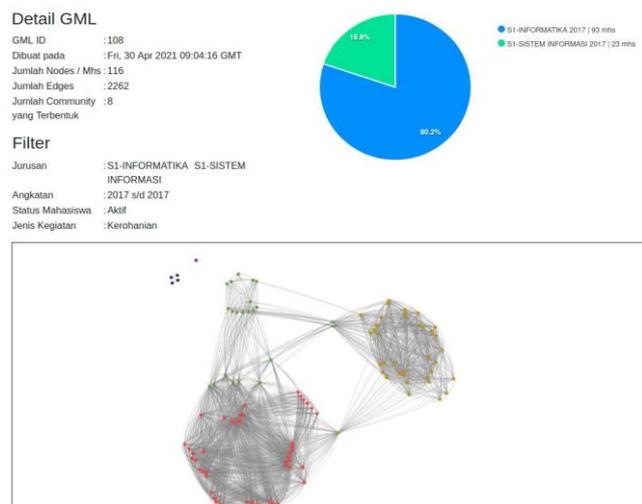
Nilai closeness centrality ini akan menunjukkan mahasiswa mana yang dianggap dapat menyalurkan informasi dengan lebih cepat ke semua mahasiswa lain, atau paling berpengaruh dan memegang informasi penting[16],[17]. Semakin berpengaruh seorang mahasiswa, biasanya semakin aktif mahasiswa tersebut dalam berkegiatan dan bersosialisasi dengan mahasiswa lainnya, dan memiliki jabatan yang tinggi seperti ketua.

Nilai closeness ini selanjutnya dihitung korelasinya dengan IPK mahasiswa tersebut dengan menggunakan pearson correlation coefficient. Karena komunitas kecil yang dihasilkan biasanya lebih dari satu dari sebuah GML, maka nilai correlation coefficient ini selanjutnya di rata-rata dengan mentransformasikan terlebih dahulu menggunakan Fisher's z-transformation. Hasil rata-rata nilai korelasi inilah yang nantinya akan diolah lebih jauh dengan melihat rata-rata nilai korelasi dari GML lain, tentu saja dengan filter dan studi kasus yang berbeda-beda.

D. Visualisasi Hasil Disjoint

Pada penelitian ini, juga dilakukan visualisasi dari hasil disjoint yang dilakukan, meskipun visualisasi merupakan hal vital dalam pembuktian hipotesa penelitian. Visualisasi akan berisi detail gml yang dihasilkan seperti jumlah nodes, jumlah edges, jumlah komunitas yang terbentuk, nilai modularity yang didapatkan, filter-filter yang berpengaruh dalam pembentukan gml, lalu akan ada visualisasi hasil akhir network beserta komunitas-komunitas yang telah ditandai dengan warna yang berbeda-beda untuk setiap nodes yang merupakan anggotanya. Berikut adalah contoh hasil visualisasi dengan menggunakan vis.js.

Visualisasi GML



Gambar. 3. Contoh visualisasi dari hasil GML dan disjoint community detection

Pada subbab-subbab berikutnya, visualisasi tidak hanya menggambarkan hasil dari disjoint community detection, tapi juga akan memvisualisasikan hasil analisa dari closeness centrality, dan average correlation coefficient.

TABEL 2
REKAP HASIL SKENARIO FILTER PER ANGGKATAN UNTUK JURUSAN ELEKTRO, INFORMATIKA, INDUSTRI, DESPRO, DKV, DAN SIB

Filter	Filter Kategori Kegiatan	Kategori Organisasi	POSITIF				NEGATIF				
			Very Strong	Strong	Moderate	Weak	No Corr	Weak	Moderate	Strong	Very Strong
2015	MIN,NAL,ORG,ROH	MIN,ORG,ROH	0.63	6.35	62.86	29.84	0.32	0	0	0	0
2015	MIN,ORG,ROH	MIN,ORG,ROH	0	9.52	37.46	40.95	11.11	0.95	0	0	0
2015	MIN,ROH	MIN,ROH	0	0	6.84	64.17	20.52	7.49	0.98	0	0
2015	ROH	ROH	0	1.61	14.84	74.84	7.1	0.32	1.29	0	0
2016	MIN,NAL,ORG,ROH	MIN,ORG,ROH	0	0	23.49	69.84	0.95	2.54	1.9	1.27	0
2016	MIN,ORG,ROH	MIN,ORG,ROH	0	0	16.51	77.78	0.63	2.22	1.59	1.27	0
2016	MIN,ROH	MIN,ROH	0	1.61	10.93	70.74	12.86	3.86	0	0	0
2016	ROH	ROH	0.33	1.63	22.8	68.08	6.84	0.33	0	0	0
2017	MIN,NAL,ORG,ROH	MIN,ORG,ROH	0	0	6.67	65.4	24.76	1.59	0	1.59	0
2017	MIN,ORG,ROH	MIN,ORG,ROH	0	0	6.51	65.8	24.43	1.63	0	1.63	0
2017	MIN,ROH	MIN,ROH	0	0.32	26.35	40.63	25.4	5.71	1.59	0	0
2017	ROH	ROH	0	0	11.75	46.67	34.29	5.71	1.59	0	0
2018	MIN,NAL,ORG,ROH	MIN,ORG,ROH	0	0	30.79	61.9	0.63	5.08	1.59	0	0
2018	MIN,ORG,ROH	MIN,ORG,ROH	0	0	27.62	70.79	0	0	1.59	0	0
2018	MIN,ROH	MIN,ROH	0	0.64	37.62	41.16	10.29	9.97	0.32	0	0
2018	ROH	ROH	1.63	1.63	1.63	2.94	53.59	35.29	0	3.27	0

* Semua angka korelasi dalam bentuk persentase (%)

** Min : minat bakat, Nal : penalaran, Roh : kerohanian, Org : organisasi

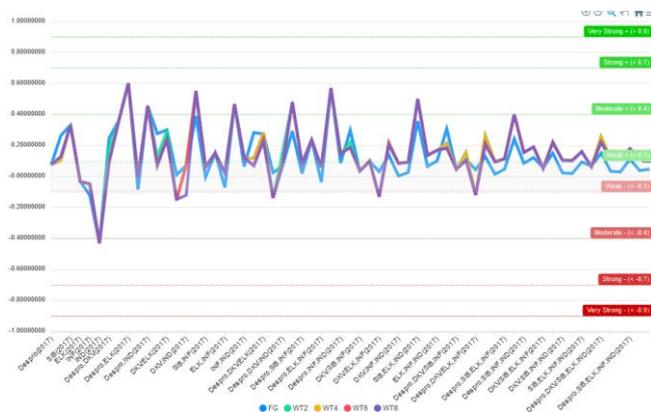
IV. HASIL EKSPERIMEN DAN UJI COBA

Pada bagian ini akan dijelaskan mengenai hasil ujicoba yang dilakukan pada berbagai macam filter dan gml. Pada semua uji coba, metode yang akan dipakai adalah fastgreedy, walktrap dengan nilai t=2,4,6, dan 8.

Beberapa skenario yang akan dibahas adalah filter per angkatan, filter per jurusan, multi-angkatan multi-jurusan, rekap top-3 mahasiswa dilihat dari IPK, rekap top-3 mahasiswa dilihat dari lama studi, dan analisa efektifitas dan performa dari kedua algoritma, serta pembahasan hasil penelitian dibandingkan dengan milik Robert dan Gail M. Ritchie

A. Skenario Filter Per Angkatan

Pada skenario ujicoba ini, gml yang terbentuk pasti hanya satu angkatan saja, misalnya hanya mahasiswa yang memiliki angkatan 2017 saja. Jurusan yang terlibat adalah elektro, informatika, industri, despro, dkv, dan sib. Jurusan lain tidak dimasukkan ke dalam uji coba dikarenakan jumlah mahasiswa yang tidak terlalu banyak. Bahkan pada beberapa jurusan seperti S2 informasi tidak menuntut mahasiswa untuk mengikuti kegiatan kemahasiswaan.



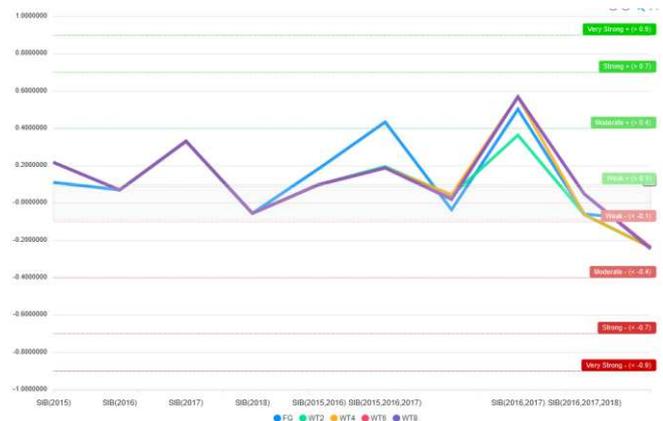
Gambar. 4. Avg. Correlation Coefficient pada angkatan 2017 kategori ROH

Pada Gambar. 4, merupakan salah satu contoh grafik average correlation, dengan filter angkatan 2017, jurusan elektro, informatika, industri, despro, dkv, dan sib. Pada

Tabel 2, dapat dilihat rekap hasil skenario beberapa sample ujicoba, untuk angkatan 2015 – 2018, dengan beberapa kombinasi kategori kegiatan, dimana dari hasil dapat disimpulkan bahwa, korelasi keseluruhan adalah : 0.16% very strong positif, 1,46% strong positif, 21.63% moderate positif, 55.75% weak positif, 14.54% tidak ada korelasi, 5.13% weak negatif, 0.78 moderate negatif, 0.56 strong negatif, dan 0% very strong negatif.

B. Skenario Filter Per Jurusan

Pada skenario uji coba ini, gml yang terbentuk pasti hanya dari satu jurusan saja, misalnya informatika. Angkatan yang akan dilibatkan adalah angkatan 2015 sampai 2018. Untuk skenario ini jurusan yang akan di uji coba hanya sib dan informatika, karena jurusan lain seperti elektro memiliki data mahasiswa yang terlalu kecil.



Gambar. 5. Avg. Correlation Coefficient pada jurusan SIB kategori ROH

Pada Gambar. 5, dapat dilihat kombinasi yang muncul menjadi lebih sedikit dibandingkan Gambar. 4. Pada Tabel 3, dapat dilihat bahwa yang diuji coba hanya jurusan informatika dengan kode jurusan 11 dan sib dengan kode jurusan 18 saja. Dari hasil percobaan ini didapatkan bahwa rekap korelasi : 0% very strong positif, 0% strong positif, 27.5% moderate positif, 42.5% weak positif, 23.5% tidak ada korelasi, 6.25% weak negatif, 0.25% moderate negatif, 0% strong negatif, dan 0% very strong negatif.

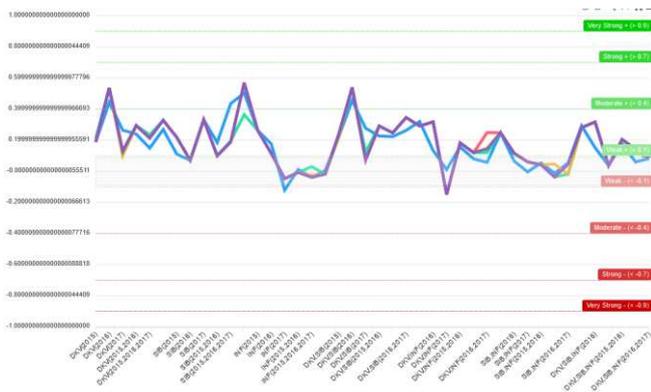
TABEL 3
REKAP HASIL SKENARIO FILTER PER JURUSAN UNTUK ANGGATAN 2015 HINGGA 2018

Filter	Filter	Kategori Kegiatan	Kategori Organisasi	POSITIF				NEGATIF				
				Very Strong	Strong	Moderate	Weak	No Corr	Weak	Moderate	Strong	Very Strong
11	MIN,NAL,ORG,ROH	MIN,ORG,ROH	MIN,ORG,ROH	0	0	74	26	0	0	0	0	0
11	MIN,ORG,ROH	MIN,ORG,ROH	MIN,ORG,ROH	0	0	68	32	0	0	0	0	0
11	MIN,ROH	MIN,ROH	MIN,ROH	0	0	28	50	18	4	0	0	0
11	ROH	ROH	ROH	0	0	0	20	76	4	0	0	0
18	MIN,NAL,ORG,ROH	MIN,ORG,ROH	MIN,ORG,ROH	0	0	10	80	10	0	0	0	0
18	MIN,ORG,ROH	MIN,ORG,ROH	MIN,ORG,ROH	0	0	18	54	12	16	0	0	0
18	MIN,ROH	MIN,ROH	MIN,ROH	0	0	12	46	22	18	2	0	0
18	ROH	ROH	ROH	0	0	10	32	50	8	0	0	0

C. Skenario Multi-Angkatan dan Multi-Jurusan

Pada skenario ujicoba ini, akan menggabungkan 6 jurusan sekaligus yaitu elektro, informatika, industri, despro, dkv, dan sib dengan 4 angkatan sekaligus yaitu 2015-2018.

Pada Gambar. 6, grafik yang dihasilkan mejadi lebih rapat karena banyaknya kombinasi, namun masih bisa dilihat bahwa hasil average correlation masih berkisar antara weak positif, dan tidak ada korelasi, dengan beberapa lonjakan average correlation ke moderate positif.



Gambar. 6. Avg. Correlation Coefficient pada kombinasi 3 jurusan (dkv,inf,sib) dengan 3 angkatan (2015-2017) pada kategori ROH

Pada Tabel 4, untuk kolom kategori kegiatan/ organisasi tidak ditampilkan karena data setiap baris sudah memuat kombinasi (1) Roh, (2) Min, Roh, (3) Min, Org, Roh, (4) Min, Org, Roh, Nal. Pada tabel ini, jumlah jurusan dan jumlah angkatan menunjukkan berapa banyak kombinasi yang dimuat dalam satu baris, sebagai contoh baris pertama

TABEL 4
REKAP HASIL SKENARIO UNTUK MULTIPLE JURUSAN DAN ANGGATAN

Filter	Jumlah Jurusan	Jumlah Angkatan	Very Strong	POSITIF				NEGATIF			
				Strong	Moderate	Weak	No Corr	Weak	Moderate	Strong	Very Strong
	2	2	0	0.11	24.56	52	19.22	3.67	0.44	0	0
	2	3	0	0	15.17	65.17	17.17	2.17	0.33	0	0
	2	4	0	0	17.67	64.67	16.33	1.33	0	0	0
	3	2	0	0	16.08	66	14.67	2.25	1	0	0
	3	3	0	0	6.5	77.5	14.5	1.5	0	0	0
	3	4	0	0	7.5	77	14.5	1	0	0	0
	4	2	0	0	9.67	74.67	14.44	0.44	0.78	0	0
	4	3	0	0	2.5	82.17	15.33	0	0	0	0
	4	4	0	0	1.67	84.33	14	0	0	0	0
	5	2	0	0	5.83	81.67	12.22	0	0.28	0	0
	5	3	0	0	2.08	83.33	14.58	0	0	0	0
	5	4	0	0	0	87.5	12.5	0	0	0	0
	6	2	0	0	0	90	10	0	0	0	0
	6	3	0	0	0	87.5	12.5	0	0	0	0
	6	4	0	0	0	90	10	0	0	0	0

dengan 2 jurusan dan 2 angkatan berarti sudah memuat kombinasi inf,sib (2015-2016), dkv,despro (2016-2017), dan seterusnya hingga semua kombinasi 2 jurusan dan 2 angkatan termasuk didalam perhitungan baris pertama.

Pada akhirnya dari skenario uji coba ini didapatkan rekap korelasi : 0% very positif, 0.01% strong positif, 11.3% moderate positif, 71.59% weak positif, 15.29% tidak ada korelasi, 1.42% weak negatif, 0.38% moderate negatif, 0% strong negatif, dan 0% very strong negatif.

D. Skenario Rekap Keseluruhan

Setelah berbagai kombinasi percobaan yang telah diujicobakan didapatkan hasil akhir dari nilai average correlation yang terbentuk antara closeness centrality dengan IPK mahasiswa untuk setiap metode yang dilakukan.

TABEL 5
REKAP HASIL KESELURUHAN BERDASARKAN METODE

Metode	Moderate Positif	Weak Positif	Tidak ada korelasi
FG	18.0	62.4	15.9
WT2	15.2	63.6	15.1
WT4	15.2	64.0	14.8
WT6	16.3	62.9	15.2
WT8	16.6	62.9	14.7

*Pada table ini, tidak ditampilkan korelasi lain seperti weak negatif, moderate negatif, dan seterusnya karena persentase yang sangat kecil (mendekati 0)

Pada Tabel 5, rekap average correlation dilihat dari setiap metode yang digunakan dalam uji coba yaitu fast greedy (FG) dan walktrap (WT) dengan nilai t=2,4,6,8. Pada ujicoba keseluruhan ini, terdapat sekitar 2527 kombinasi filter yang mencampur jurusan, angkatan, dan kategori kegiatan/organisasi. Dari hasil yang ditampilkan, terlihat bahwa hampir semua metode mengeluarkan hasil weak

positif untuk korelasi antara closeness centrality mahasiswa dengan ipknya, diikuti dengan moderate positif dan tidak ada korelasi.

E. Rekap Top-3 Dilihat dari IPK

Pada subbab ini akan dijelaskan bagaimana informasi hasil yang sudah didapatkan sebelumnya, diolah lebih jauh untuk kontribusi yang lebih jauh. Dengan mengambil data mahasiswa yang paling sering muncul sebagai top, dalam hal ini kita sebut saja sebagai top-3 mahasiswa dengan nilai closeness centrality tertinggi dalam sebuah komunitas hasil disjoint community detection, maka bisa didapatkan beberapa informasi seperti mahasiswa mana saja yang dianggap sebagai mahasiswa yang paling “penting” dalam berbagai komunitas yang ada, bagaimana performa mahasiswa-mahasiswa top ini dalam bidang akademik seperti IPK.

Kedua algoritma menghasilkan sekitar 15 ribu sampai 20 ribu kemunculan data mahasiswa yang dianggap sebagai top-3 closeness centrality dari semua GML yang dihasilkan. Dari hasil rekap, dapat dilihat bahwa rata-rata 30-32% mahasiswa memiliki IPK 3-3.5 dengan predikat sangat memuaskan, 22-24% mahasiswa memiliki IPK 3.7-3.9 dengan predikat magna cumlaude, 19-21% mahasiswa memiliki IPK 2.5-3 dengan predikat memuaskan, 15-16% mahasiswa memiliki IPK 3.5-3.7 dengan predikat cumlaude, dan sisanya, seperti IPK kurang dari 2.5 atau IPK diatas 3.7

Dari hasil tersebut, maka pada studi kasus di ISTTS, tidak benar bila ada anggapan bahwa mahasiswa yang aktif berkegiatan, maka akan memiliki nilai IPK yang jelek, sebaliknya justru sebagian besar mahasiswa ini memiliki IPK yang bisa dibilang cukup baik.

F. Rekap Top-3 Dilihat dari Masa Studi

Pada subbab ini, akan dibahas mengenai rekap top-3 mahasiswa dilihat dari masa studi yang telah ditempuh oleh mahasiswa. Untuk jumlah data mahasiswa top-3 masih sama dengan rekap top-3 mahasiswa dilihat dari IPK yang dicapai, seperti yang dijelaskan pada subbab sebelumnya.

Hasil yang didapatkan adalah 45%-46% mahasiswa belum lulus, 34%-36% mahasiswa lulus tepat waktu, 16-17% mahasiswa lulus, namun kelebihan satu tahun dari waktu seharusnya, sisanya, adalah mahasiswa yang lulus tapi lebih dari 2 tahun dari waktu seharusnya.

Dari hasil ini dapat disimpulkan bahwa, pada studi kasus ISTTS, tidak benar kalau ada pernyataan yang mengatakan bahwa mahasiswa yang aktif berkegiatan, tidak lulus-lulus / lulusnya terlambat. Dari dua buah contoh rekap top-3 mahasiswa dari lama studi dan IPK, dapat dilihat bahwa penelitian dalam bidang disjoint community detection ini, bisa menghasilkan banyak sekali deskriptif model untuk menghasilkan sejumlah informasi baru yang sebelumnya belum diketahui.

G. Analisa Performa dan Hasil

Pada subbab ini akan sedikit dijelaskan mengenai analisa performa dari fastgreedy dan walktrap. Pada penelitian ini, tidak akan dijelaskan mengenai bagaimana performa kecepatan dan nilai modularity yang dihasilkan dari kedua algoritma ini, dalam menghandle network dalam ukuran besar, karena hal ini telah disampaikan oleh Pascal et al pada papernya[7].

Pada penelitiannya, Pascal menunjukkan perbandingan fast greedy dengan walktrap dengan nilai $t=2$ dan $t=5$, dimana kedua algoritma ini menghasilkan nilai modularity / time dalam detik yang hampir sama pada network kecil seperti karate club, dan nilai modularity yang lebih kecil dari fast greedy pada network www, namun dengan peningkatan time waktu pada walktrap $t=2$ dibandingkan dengan fast greedy.

Pada penelitian ini, dilihat dari subbab-subbab sebelumnya dapat dilihat bahwa sekitar 60% menyatakan ada relasi weak positif antara kegiatan kemahasiswaan dengan IPK yang diraih oleh mahasiswa. Hasil ini sejalan dengan dua paper rujukan yaitu milik Robert dan milik Gail M. Ritchie, sehingga dapat ditarik kesimpulan untuk menjawab research question pada penelitian ini, yaitu ada pengaruh meskipun secara weak antara keaktifan mahasiswa dengan IPK yang diraih. Selain itu, didapatkan bahwa semakin aktif mahasiswa tersebut, bahkan bila mahasiswa tersebut menjadi mahasiswa penting atau top-3, maka IPK yang diraih juga cukup baik dan lama studi mereka sebagian besar adalah lulus tepat waktu.

V. KESIMPULAN

Dari hasil penelitian yang telah dilakukan, dapat diambil beberapa kesimpulan. Hubungan antara nilai closeness centrality seorang mahasiswa ISTTS yang menandakan mahasiswa tersebut adalah pusat dari sebuah komunitas, dengan ipk, didapatkan mahasiswa tersebut cenderung weak positif, hal ini dapat dilihat dari perhitungan korelasi secara keseluruhan menyatakan 62 - 63% weak positif dengan diikuti 16-18% moderate positif, dan 14-16% tidak berkorelasi sama sekali.

Hasil komunitas yang dibentuk oleh ketiga jenis metode ini pun hampir sama dalam pembentukan komunitasnya, terlihat dari hasil average closenessnya yang tidak terlalu sering berbeda terlalu jauh, terutama untuk walktrap dengan nilai $t=6$ dan 8.

Data kemahasiswaan, terutama mahasiswa ISTTS tergolong cukup kecil dimana terdapat trend banyak sekali mahasiswa yang terlibat dalam suatu kegiatan terus menerus atau malah tidak terlibat kegiatan apapun sama sekali, ini terbukti dari bentuk graph/gml yang dihasilkan selalu berbentuk dense graph, sekaligus disconnected graph terhadap beberapa mahasiswa yang enggan ikut kegiatan apapun.

Beberapa masalah yang muncul dalam penelitian ini adalah algoritma walktrap terutama yang didevelop oleh igraph sering menghasilkan incomplete dendogram, sehingga tidak bisa didapatkan hasil komunitasnya/ cluster. Selain itu komunitas yang terbentuk pun sering berupa disconnected graph yang mempersulit perhitungan closeness centrality.

Beberapa saran untuk penelitian selanjutnya adalah menggunakan function `fix_dendogram` untuk membenarkan dendogram milik walktrap, dan mengganti closeness centrality dengan harmonic centrality. Penelitian juga dapat dikembangkan untuk menjadi deskriptif modelling untuk mendapatkan informasi baru yang sebelumnya belum diketahui, seperti contohnya pada penelitian ini dilakukan analisa lebih jauh mengenai rekap top-3 mahasiswa dan

pengaruhnya pada IPK dan lama studi.

DAFTAR PUSTAKA

- [1] R. Freeman, "The relationship between extracurricular activities and academic achievement," 2017.
- [2] G. M. Ritchie, "The Impact of Academic Co-Curricular Activity Participation on Academic Achievement: A Study of Catholic High School Students," 2018.
- [3] S. Fortunato, "Community detection in graphs," *Phys. Rep.*, vol. 486, no. 3–5, pp. 75–174, 2010.
- [4] M. E. J. Newman, "Fast algorithm for detecting community structure in networks," *Phys. Rev. E*, vol. 69, no. 6, p. 66133, 2004.
- [5] M. E. J. Newman, "Detecting community structure in networks," *Eur. Phys. J. B*, vol. 38, no. 2, pp. 321–330, 2004.
- [6] S. Paul and Y. Chen, "Null models and modularity based community detection in multi-layer networks," *arXiv Prepr. arXiv1608.00623*, 2016.
- [7] P. Pons and M. Latapy, "Computing communities in large networks using random walks," in *International symposium on computer and information sciences*, 2005, pp. 284–293.
- [8] J. Zhang and Y. Luo, "Degree centrality, betweenness centrality, and closeness centrality in social network," in *Proceedings of the 2017 2nd International Conference on Modelling, Simulation and Applied Mathematics (MSAM2017)*, 2017, vol. 132, pp. 300–303.
- [9] T. Opsahl, F. Agneessens, and J. Skvoretz, "Node centrality in weighted networks: Generalizing degree and shortest paths," *Soc. Networks*, vol. 32, no. 3, pp. 245–251, 2010.
- [10] P. Schober, C. Boer, and L. A. Schwarte, "Correlation coefficients: appropriate use and interpretation," *Anesth. & Analg.*, vol. 126, no. 5, pp. 1763–1768, 2018.
- [11] D. M. Corey, W. P. Dunlap, and M. J. Burke, "Averaging correlations: Expected values and bias in combined Pearson rs and Fisher's z transformations," *J. Gen. Psychol.*, vol. 125, no. 3, pp. 245–261, 1998.
- [12] A. A. Hagberg, D. A. Schult, and P. J. Swart, "Exploring Network Structure, Dynamics, and Function using NetworkX," in *Proceedings of the 7th Python in Science Conference*, 2008, pp. 11–15.
- [13] G. Csardi and T. Nepusz, "The igraph software package for complex network research," *InterJournal*, vol. Complex Sy, p. 1695, 2006.
- [14] G. A. Pavlopoulos, D. Paez-Espino, N. C. Kyrpides, and I. Iliopoulos, "Empirical comparison of visualization tools for larger-scale network analysis," *Adv. Bioinformatics*, vol. 2017, 2017.
- [15] G. Perrone, J. Unpingco, and H. Lu, "Network visualizations with Pyvis and VisJS," *CoRR*, vol. abs/2006.0, 2020.
- [16] V. E. Krebs, "Mapping networks of terrorist cells," *Connections*, vol. 24, no. 3, pp. 43–52, 2002.
- [17] S. P. Borgatti, "Centrality and network flow," *Soc. Networks*, vol. 27, no. 1, pp. 55–71, 2005.