

Perbandingan Metode Single Linkage dan Fuzzy C Means Untuk Pengelompokan Trafik Internet

Auliya Burhanuddin¹, Ema Utami², Eko Pramono³

¹Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Magelang,

²Magister Teknologi Informatika, Stimik Amikom Yogyakarta, Yogyakarta

³Magister Teknologi Informatika, Stimik Amikom Yogyakarta, Yogyakarta

email: ¹auliyaburhanuddin@gmail.com, ²Emma@nrar.net, ³Ekopramonoid@yahoo.com

Abstrak

Penelitian dengan judul “Perbandingan metode Single Linkage dan Fuzzy C Means untuk pengelompokan trafik internet” bertujuan mengetahui kluster pengelompokan bandwidth internet menggunakan metode Single Linkage dan Fuzzy C Means. Batasan penelitian adalah pengelompokan dengan metode clustering Single Linkage dan Fuzzy C Means dan akan dibandingkan untuk mengetahui kualitas kluster yang lebih baik menggunakan metode purity. Data yang digunakan adalah data upload, download, dan jumlah login tiap user dari 1 Agustus 2016 s/d 28 Januari 2017. Single Linkage merupakan hierarchi clustering, data dikelompokkan tiap data terdekat sehingga diakhir didapatkan kelompok tertentu. Fuzzy C Means merupakan partitional clustering dimana dari data akan dicari pusat kluster dan data yang lebih dekat ke pusat kluster akan menjadi anggota kelompok tersebut. Kualitas kluster diuji dengan metode Purity, kluster yang mempunyai nilai purity mendekati 1 adalah kluster yang mempunyai kualitas yang lebih baik. Perbandingan dua metode didapatkan metode mana yang menghasilkan kelompok dengan kualitasnya lebih baik dan kecepatan pengelompokannya yang lebih cepat. Jumlah data uji 50, 100, 150, 200, 250, 300, 350, dan 400. Hasilnya didapatkan Single Linkage mempunyai kualitas kluster lebih baik karena nilai puritinya lebih tinggi. Tetapi kecepatan pengelompokannya Fuzzy C Means lebih cepat. Kelompok trafik internet yang didapatkan adalah kelompok pengguna internet paling besar 880 user, kelompok sedang 12 user, dan kelompok kecil 2106 user. Dari 8 sampel data uji didapatkan kualitas kluster single linkage lebih baik dibandingkan fuzzy c means, tetapi kecepatan pengelompokan fuzzy c means lebih cepat dibandingkan single linkage. Semakin lama proses pengelompokan akan didapatkan kualitas kluster yang lebih baik.

Kata Kunci: clustering, Fuzzy C Means, Single Linkage, Purity

I. PENDAHULUAN

Dalam kehidupan kita sehari-hari disetiap sudut kota, internet sudah beralih menjadi kebutuhan yang wajib ada, terutama di suatu lembaga pendidikan. Dengan internet kita dapat terhubung dengan siapa saja tanpa terhalang oleh jarak. Karena internet adalah layanan jaringan dari komputer yang saling terhubung menggunakan sebuah perangkat jaringan agar bisa terkoneksi ke internet.

Universitas Muhammadiyah Magelang (UMMagelang) adalah salah satu perguruan tinggi di Magelang yang memiliki 6 fakultas dan 16 program studi dan mempunyai jumlah mahasiswa aktif pada semester ganjil tahun 2015/2016 sebanyak 4777 mahasiswa dan pada semester genap tahun 2015/2016 sebanyak 4264 mahasiswa. Dalam mendukung aktivitas perkuliahan penggunaan internet di kampus sangat diperlukan untuk dapat mendukung aktivitas perkuliahan. Sehingga untuk memperlancar akses internet, maka saat ini bandwidth yang dimiliki dibagi rata ke tiap-tiap fakultas. Dengan besarnya bandwidth yang dimiliki diharapkan dapat digunakan secara

dengan optimal. Tetapi hal tersebut masih kurang adil, karena pada waktu-waktu tertentu ada perbedaan jumlah pengakses internet di tiap-tiap fakultas dan banyak bandwidth masih dirasa kurang karena pada jam-jam tertentu ada fakultas yang trafik internetnya sangat padat karena pengakses internet sangat banyak yang mengakibatkan kecepatan internetnya menjadi lambat. Tetapi pada jam yang sama ada fakultas yang trafik internetnya tidak padat. Untuk mengatasi hal tersebut maka perlu dibuat pembagian bandwidth dengan mengelompokkan penggunaannya agar pembagian bandwidthnya dibagi dengan tepat dengan cara mengelompokkan/mengkluster.

II. PENELITIAN YANG TERKAIT

Pada penelitian sebelumnya telah dilakukan klustering dan analisa trafik internet menggunakan Fuzzy C-Means dengan ekstraksi fitur data. Pada penelitian tersebut menggunakan metode klustering yaitu metode Fuzzy C-Means dimana pada proses awal sebelum klusterisasi data penggunaan bandwidth internet yang ada dalam satu periode

akan dikumpulkan untuk menjadi inputan pada algoritma Fuzzy C-Mean untuk dilakukan pembagian kluster terhadap penggunaan bandwidth yang ada berdasarkan aplikasi yang digunakan dan pemakai jaringan internet [1].

Metode kluster tidak hanya metode Fuzzy C-Means, tetapi masih ada metode lain salah satunya adalah metode K-Means. Dari dua metode tersebut telah dilakukan penelitian perbandingannya. Hasil perbandingan dua metode tersebut telah dilakukan dengan judul “Implementation of the Fuzzy C-Means Clustering Algorithm in Meteorological Data”. Dalam penelitian tersebut, algoritma Fuzzy C Means memiliki perhitungan berulang lebih cepat dari pada algoritma K-Means. Metode lain dalam clustering adalah metode single linkage Metode ini menghasilkan clustering hirarkis (dendogram). Ada algoritma yang tidak memperhatikan ketidakpastian. Dalam metode ini diperoleh model statistik untuk estimasi dendogram dengan memperhatikan ketidakpastian (melalui noise dan corrupt) jarak antar titik data. Hasil cluster yang didapatkan berupa beberapa kluster yang unsupervised [2].

metode Single Linkage, metode tersebut dibandingkan dengan metode K-Means yang diteliti oleh Rendy Handoyo, R Rumani M, dan Surya Michrandi Nasution dengan judul “Perbandingan metode Clustering menggunakan metode Single Linkage dan K-Means pada pengelompokan Dokumen”. Dalam penelitian tersebut menjelaskan bahwa metode Single Linkage memiliki performansi yang lebih baik dibandingkan dengan metode K-Means [3].

Dalam Penelitian dengan Judul “Pengelompokan Jenis Kupu-Kupu Menggunakan Fitur Ekstraksi GLCM dan Algoritma K-Means”. Dalam penelitian tersebut Metode *Purity* digunakan sebagai evaluasi kinerja algoritma clustering. *Purity* akan menghasilkan persentase akurasi kinerja clustering [4].

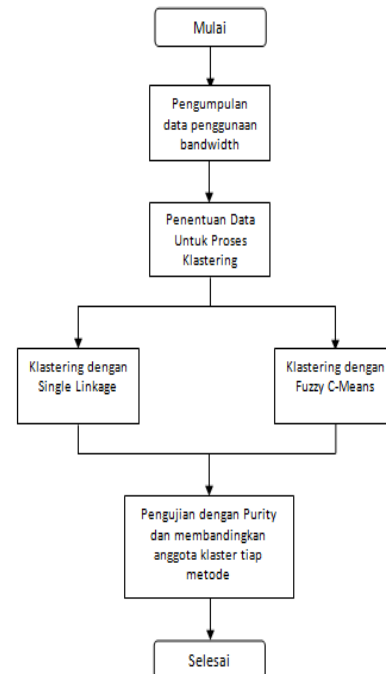
Perbandingan lain yang sudah diteliti adalah perbandingan antara metode kluster Single Linkage dan metode K-Means. Menurut Handoyo, R at al bahwa Pada penelitian ini dilakukan pengujian dengan menggunakan untuk mengukur kualitas cluster adalah *Silhouette Coefficient* dan *Purity*. Hasil dari penelitian ini adalah metode Single Linkage memiliki performansi yang lebih baik dibandingkan dengan metode K-Means. Nilai *Silhouette Coefficient* Single Linkage selalu lebih unggul dibandingkan dengan K-Means. Pertambahan jumlah dokumen membuat

nilai *Silhouette Coefficient* Single Linkage semakin kecil sedangkan K-Means terkadang menghasilkan nilai yang negatif. Untuk nilai *Purity*, Single Linkage selalu bernilai 1 sedangkan K-Means tidak pernah bernilai 1. Hasil penambahan jumlah cluster dan jumlah dokumen memberikan pengaruh terhadap nilai *Silhouette Coefficient* dan *Purity*. Hal ini berarti single linkage selalu menghasilkan dokumen yang sama, sedangkan K-Means masih bercampur dengan dokumen lain [5].

III. METODE PENELITIAN

Alur penelitian yang dilakukan dalam penelitian ini dapat dilihat pada flowchart berikut:

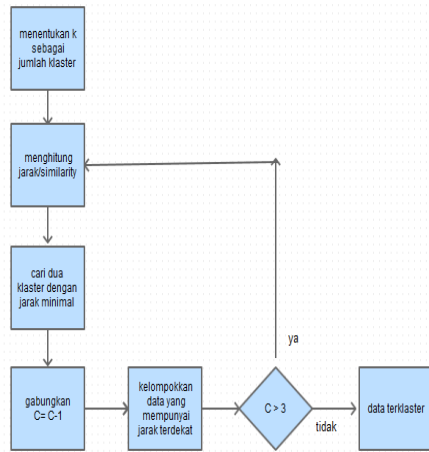
A. Flowchart Penelitian



Gambar 1. Flowchart Utama Penelitian

Perbandingan yang akan dilakukan adalah perbandingan klustering antara metode Single Linkage dan metode Fuzzy C Means. Untuk pengujian akan dilakukan pengujian kualitas dengan metode *Purity* dan membandingkan anggota klasternya.

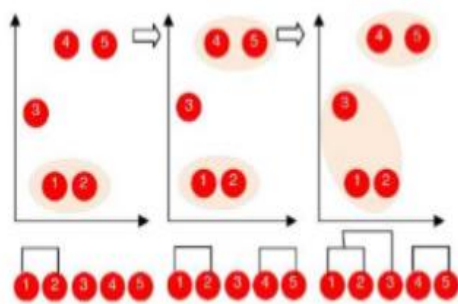
B. Flowchart metode Klastering Single Linkage



Gambar 2. Flowchart metode Single Linkage

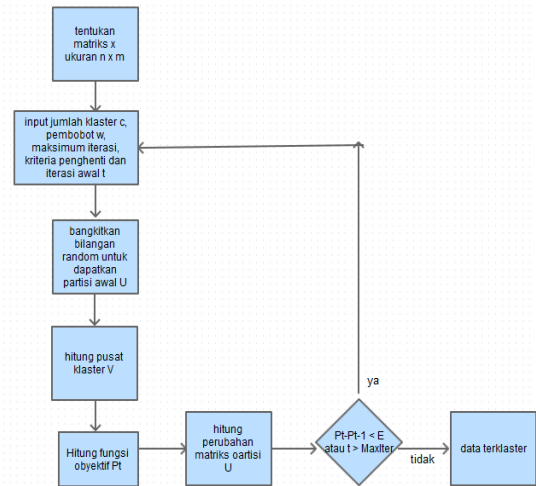
Langkah-langkah metode single linkage adalah sebagai berikut:

1. Menentukan k sebagai jumlah cluster yang ingin dibentuk.
2. Setiap data dianggap sebagai cluster. Kalau n = jumlah data dan c = jumlah cluster, berarti ada c = n.
3. Menghitung jarak/similarity/dissimilarity antar cluster.
4. Cari dua cluster yang mempunyai jarak antar cluster yang minimal dan gabungkan (c=c-1). Setelah semua jarak diketahui, selanjutnya dikelompokkan data yang memiliki jarak terdekat.
5. Jika c>3, kembali ke langkah 3 [3].



Gambar 3. Langkah-langkah metode single linkage

C. Flowchart metode Klustering Fuzzy C Means



Gambar 4. Flowchart metode Fuzzy C Means

Langkah-langkah metode Fuzzy C Means adalah sebagai berikut:

1. Tentukan:
 - a. Matriks X berukuran n x m, dengan n = jumlah data yang akan diklaster, dan m = jumlah variabel (kriteria).
 - b. Jumlah kluster yang akan dibentuk = C (≥ 2).
 - c. Pangkat (pembobot) = w (> 1)
 - d. Maksimum iterasi (MaxIter)
 - e. Kriteria penghentian = ξ (nilai positif yang sangat kecil)
 - f. Iterasi awal, t = 1, dan Δ = 1;
2. Bentuk matriks partisi awal, U⁰, sebagai berikut:

$$U = \begin{bmatrix} U_{11}(X_{11}) & U_{12}(X_{12}) & \dots & U_{1k}(X_{1k}) \\ U_{21}(X_{21}) & U_{22}(X_{22}) & \dots & U_{2k}(X_{2k}) \\ \vdots & \vdots & \ddots & \vdots \\ U_{i1}(X_{i1}) & U_{i2}(X_{i2}) & \dots & U_{ik}(X_{ik}) \end{bmatrix} \quad (1)$$

(matriks partisi awal bisaanya dipilih secara acak)

3. Hitung pusat kluster, V, untuk setiap kluster.

$$V_{ij} = \frac{\sum_{k=1}^n (\mu_{ik})^w \cdot X_{kj}}{\sum_{k=1}^n (\mu_{ik})^w} \quad (2)$$

4. Perbaiki derajat keanggotaan setiap data pada setiap kluster (perbaiki matriks partisi), sebagai berikut:

$$\mu_{ik} = \left[\sum_{j=1}^c \left(\frac{d_{ik}}{d_{jk}} \right)^{2/(w-1)} \right]^{-1} \quad (3)$$

Dengan:

$$d_{ik} = d(x_k - v_i) = \left[\sum_{j=1}^m (X_{kj} - V_{ij}) \right]^{1/2} \quad (4)$$

5. Tentukan kriteria berhenti, yaitu perubahan matriks partisi pada iterasi sekarang dengan iterasi sebelumnya, sebagai berikut:

$$\Delta = \|U^t - U^{t-1}\| \tag{5}$$

Apabila $\Delta \leq \epsilon$, maka iterasi dihentikan, namun apabila $\Delta > \epsilon$, maka naikan iterasi ($t=t+1$) dan kembali ke langkah-3.

Pencarian nilai D dapat dilakukan dengan mengambil elemen terbesar dari nilai mutlak selisih antara $\mu_{ik}(t)$ dengan $\mu_{ik}(t - 1)$.

D. Metode Pengujian Purity

Purity (kemurnian) suatu cluster direpresentasikan sebagai anggota cluster yang banyak sesuai (cocok) di suatu kelas. Purity dapat dihitung dengan rumus berikut :

$$Purity(j) = \frac{1}{n_j} \max(n_{ij}) \tag{6}$$

Total nilai Purity dapat dihitung dengan rumus berikut (Hndoyo, dkk, 2014):

$$Purity = \sum_{j=0}^n \frac{n_j}{n} Purity(j) \tag{7}$$

IV. HASIL DAN PEMBAHASAN

Dalam implemantasi metode Fuzzy C Means dan Single Linkage untuk mengetahui kualitas klaster dan kecepatan prosesnya maka perlu diimpelentasikan dengan beberapa jumlah data. Data yang akan diimplementasikan didapatkan dari hasil record dengan mikrotik dalam waktu 6 bulan (1 Agustus 2016 s/d 28 Januari 2017) yaitu sebanyak 3000 data (Lampiran 1).Tetapi untuk mengetahui perbedaan antara metode Fuzzy C Means dengan Single Linkage, maka banyaknya jumlah data yang digunakan adalah 50, 100, 150, 200, 250, 300, 350, dan 400 data. Berikut implementasi dengan beberapa data untuk tiap metode klaster:

Hasil pengelompokkan dengan 50 data dengan Fuzzy C Means adalah sebagai berikut:

Tabel 1. Hasil klastering 50 data dengan Fuzzy C Means

Kelompok	Jumlah	Anggota
Kelompok 1	6	1506020047, 1404010048, 1404010052, 1305040067, 1406030013, 1503050034.
Kelompok 2	31	1203050067, 1303050088, 1106030031, 1105040095, 1001020033, 1301010029, 1506030052, 1506030005, 27207033, 1304010031, 1505020024, 1506040019, 1205040004, 1205040033, 1503010025, 1105040048, 1203050015, 1503010027, 1501020223, 905040046, 1305040089, 1501010204, 1305040103, 1104050017, 1501020007, 1503050134, 1205040086, 1203050134, 1505030018, 1501010225, 1505040026.
Kelompok 3	13	1305040061, 1304010032, 1405030013, 1305040108, 120504095, 1505030037, 1105010005, 1205010011, 1305040080, 1505040105, addin, 1306030026, 1504040008.

Hasil pengelompokkan dengan 50 data dengan Single Linkage adalah sebagai berikut:

Tabel 2. Hasil klastering 50 data dengan single linkage

Kelompok	Jumlah	Anggota
Kelompok 1	3	1205040095, 1105010005, 1105040048.
Kelompok 2	27	1106030031, 1001020033, 1301010029, 1506030052, 1404010048, 1304010032, 27207033, 1405030013, 1505030037, 1505020024, 1205040033, 1503010025, 1404010052, 1203050015, 1305010011, 1503010027, 1501020223, 1305040080, 1305040103, 1503050134, 1505040105, 1205040086, 1406030013, 1501010225, 1505040026, 1503050034, 1504040008.
Kelompok 3	20	1203050067, 1305040061, 1303050088, 1105040095, 1506020047, 1506030005, 1304010031, 1305040108, 1506040019, 1205040004, 905040046, 1305040089, 1501010204, 1104050017, 1501020007, 1305040067, 1203050134, addin, 1505030018, 1306030026

Dari proses implementasi dengan jumlah data 50, 100, 150, 200, 250, 350, dan 400 didapatkan hasil jumlah tiap klasternya sebagai berikut:

Tabel 3. Jumlah anggota tiap klaster dengan Fuzzy C Means dan Single Linkage

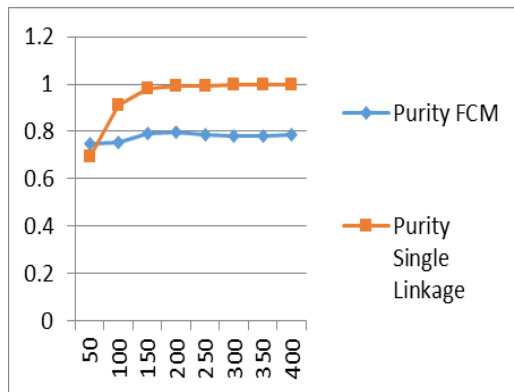
No	Banyak Data	Metode Fuzzy C Means			Metode Single Linkage		
		Kel 1	Kel 2	Kel 3	Kel 1	Kel 2	Kel 3
1	50	6	31	13	3	27	20
2	100	8	63	29	0	87	13
3	150	8	101	41	1	146	3
4	200	2	139	59	1	198	1
5	250	5	170	75	1	248	1
6	300	7	202	91	0	299	1
7	350	9	234	107	1	348	1
8	400	8	273	119	0	398	2

Hasil Purity dari metode kluster Fuzzy C Means dan Single Linkage adalah sebagai berikut:

Tabel 4.
Purity dari Fuzzy C Means dan Single Linkage

No	Banyak Data	Purity Fuzzy C Means	Purity Single Linkage
1	50	0,7467	0,69333
2	100	0,7533	0,91333
3	150	0,7911	0,98222
4	200	0,7967	0,99333
5	250	0,7867	0,99467
6	300	0,7822	0,99778
7	350	0,7790	0,99619
8	400	0,7883	0,99667

Untuk lebih jelasnya dapat dilihat pada grafik berikut:



Gambar 5. Hasil Perbandingan Purity

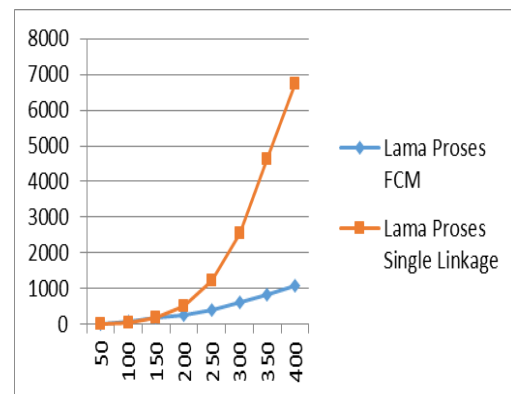
Dari hasil purity dan grafik tersebut dapat dilihat untuk data yang sedikit nilai purity Fuzzy C Means lebih besar dibandingkan Single Linkage, akan tetapi dengan bertambahnya jumlah data maka nilai purity Single Linkage unggul dibandingkan purity Fuzzy C Means.

Waktu proses clustering dari proses clustering dengan metode Fuzzy C Means dan Single Linkage untuk beberapa data uji didapatkan data waktu proses sebagai berikut:

Tabel 5. Tabel Waktu Proses Clustering

No	Banyak Data	Fuzzy C Means (sekon)	Single Linkage (sekon)
1	50	16.4014	4,2988
2	100	68.6432	39,9583
3	150	208.009	174,8477
4	200	269.6305	516,8229
5	250	418.8463	1220,9716
6	300	606.53625	2554,7705
7	350	827.2266	4637,2779
8	400	1083.42715	6736,8475

Untuk lebih jelasnya dapat dilihat di grafik gbr 6.



Gambar 6. Grafik waktu proses clustering

Dari grafik terlihat bahwa semakin bertambahnya data uji maka untuk metode Single Linkage lama proses semakin lama, sedangkan lama proses metode Fuzzy C Means juga semakin bertambah tetapi tidak signifikan. Hal ini terjadi karena dalam metode Single Linkage, tidak dapat ditentukan maksimum iterasinya. Sehingga proses clustering berlangsung jika proses perbandingan satu persatu data selesai hingga kelompok kluster terpenuhi. Sedangkan pada metode Fuzzy C Means, kita dapat menentukan maksimum iterasi yang akan dilakukan jika syarat berhenti iterasi tidak terpenuhi.

V. KESIMPULAN

1. Dari 8 sampel data uji (50, 100, 150, 200, 250, 300, 350, dan 400 data) didapatkan kualitas kluster Single Linkage lebih baik dibandingkan dengan Fuzzy C Means.
2. Dari 8 sampel data uji didapatkan waktu proses metode Single Linkage lebih lama dibandingkan dengan metode Fuzzy C Means.

3. Semakin banyak iterasi yang dilakukan maka kualitas cluster yang didapatkan akan semakin baik.

UCAPAN TERIMA KASIH

1. Terima Kasih banyak kepada Ibu Prof. Dr. Ema Utami, S.Si., M.Kom yang telah membimbing dan memotivasi dalam penelitian ini, semoga ilmu yang diberikan akan selalu tercurah dan menjadi amal jariah.
2. Terima kasih banyak kepada Bpk Eko Pramono, S.Si., M.T yang telah membimbing dan memotivasi dalam penelitian ini, semoga ilmu yang diberikan akan selalu tercurah dan menjadi amal jariah.
3. Terima kasih penulis kepada semua pihak yang membantu ataupun memberikan dukungan yang tidak dapat disebutkan satu persatu

DAFTAR PUSTAKA

- [1] Suryaputra P, A; Samopa, F; Hindayanto, BC, Klasterisasi Analisis Traffik Internet Menggunakan *Fuzzy C-Means* dengan Ekstraksi Fitur Data, In Jurnal Informatika ISSN 1411-0105 Vol. 12, No. 1, Mei 2014, pp 33-39.
- [2] Lu, Y; Ma, T; Yin, C; Xie, X; Tian, W; Zhong, SM, Implementation of the Fuzzy C-Means Clustering Algorithm in Meteorological Data, In International Journal of Database Theory and Application, ISSN: 2005-4270 IJDTA Vol.6, No 6 (2013), pp. 1-18.
- [3] Handoyo, R; Rumani M, R; Nasution, SM, Perbandingan Metode *Clustering* Menggunakan Metode *Single Linkage* dan *K – Means* Pada Pengelompokan Dokumen. In JSM STIMIK Mikroskil ISSN. 1412-0100 ISSN: 1412-0100 Vol 15, No 2, Oktober 2014, pp. 73-82.
- [4] Kusuma, T, Pengelompokkan Jenis Kupu-Kupu Menggunakan Fitur Ekstraksi GLCM dan Algoritma K-Means, Teknik Informatika, Universitas Dian Nuswantoro, Semarang.
- [5] Zhu, D; Guralnik, D; Wang, X; Li, X; Moran, B, Statistical Estimation for Single Linkage Hierarchical Clustering, The 5th Annual IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems June 8-12, 2015, Shenyang, China.