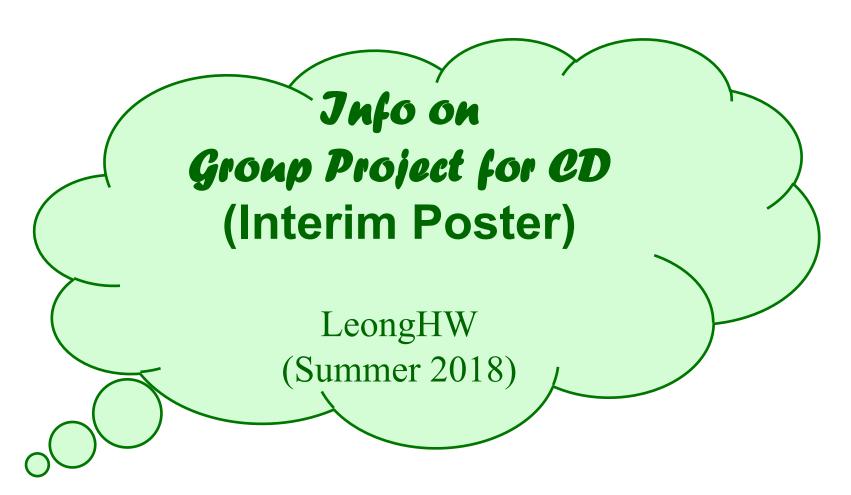
CD Project



School of Computing



Project Milestones (2018)

Team Project: (Milestones)

M1: Project proposal

[23-July]

M2: Interim Poster Report [26-July]

M3: Poster Showcase

[01-Aug]

Milestone M2: Interim Poster...

- ☐ Interim Poster reports progress in your project.
- **■** Expectations:
 - * To have got project data;
 - * And maybe... with test data
 - * To have computed dissimilarity matrix
 - * To have determined cutoff, form Graph
 - Computed communities
 - Maybe start to analyze communities

Preparing your Interim Poster...

□ Download template poster from IVLE > File > Project

- ☐ Use it to prepare your Interim Poster
- **☐** Two purposes:
 - Summary of your progress
 - *And preparation for your Final Poster...

Submitting your Interim Poster

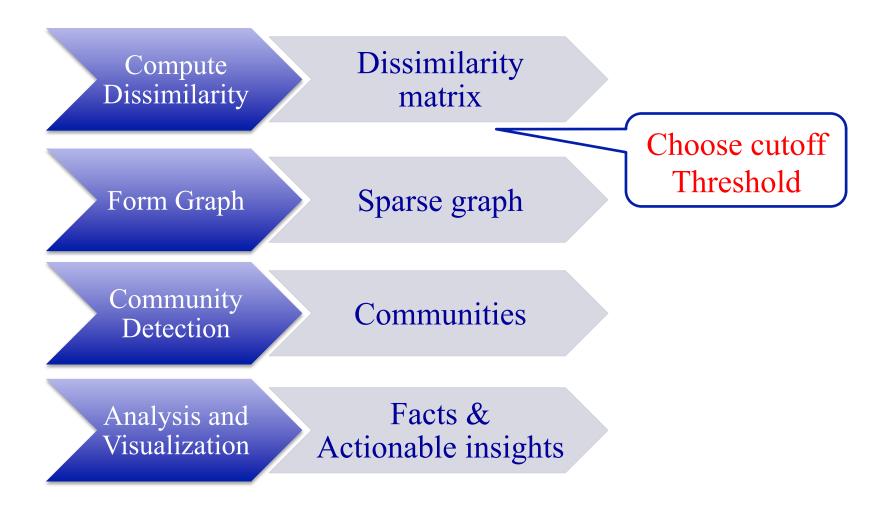
- □ Call your poster G03-Interim.pptx
- ☐ Create pdf version G03-Interim.pdf
- □ Submit *both* files to:

IVLE > Project > Group 03 > Files

□ Deadline: 26-Thu, 11:59am

☐ Any extra details goes to separate ppt file.

CD-Lab Workflow...

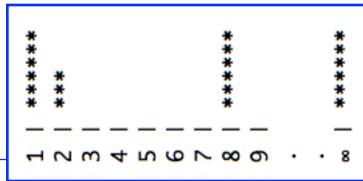


Choosing dissimilarity cut-offs

Dissimilarity matrix

	1	2	3	4	5	6	7
1	0	8	1	8	1	8	∞
2	8	0	8	1	8	2	1
3	1	8	0	∞	1	2	8
4	∞	1	∞	0	8	∞	1
5	1	8	1	8	0	2	∞
6	8	2	2	8	2	0	8
7	8	1	8	1	∞	8	0

A sufficiently large dissimilarity, namely $(\geq T)$ can be considered as ∞ .

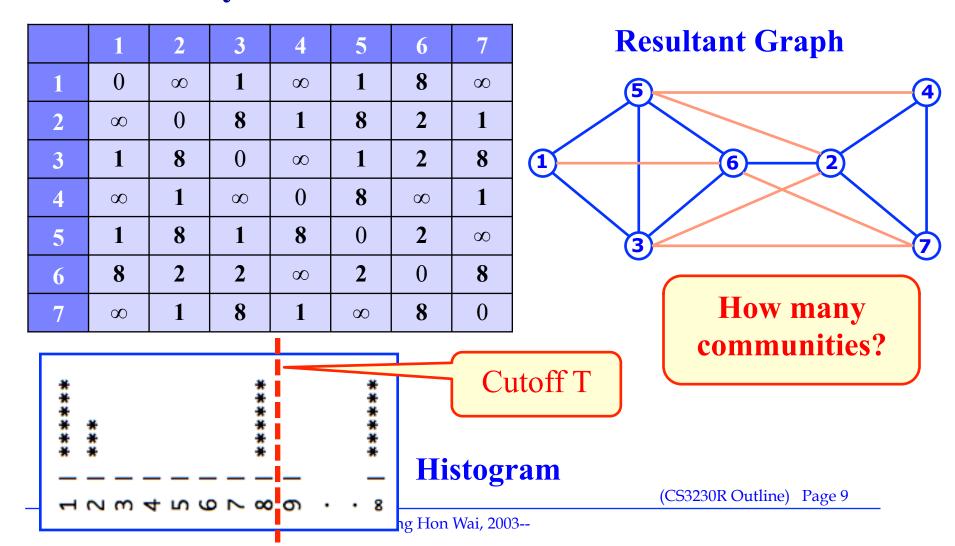


Histogram

(CS3230R Outline) Page 8

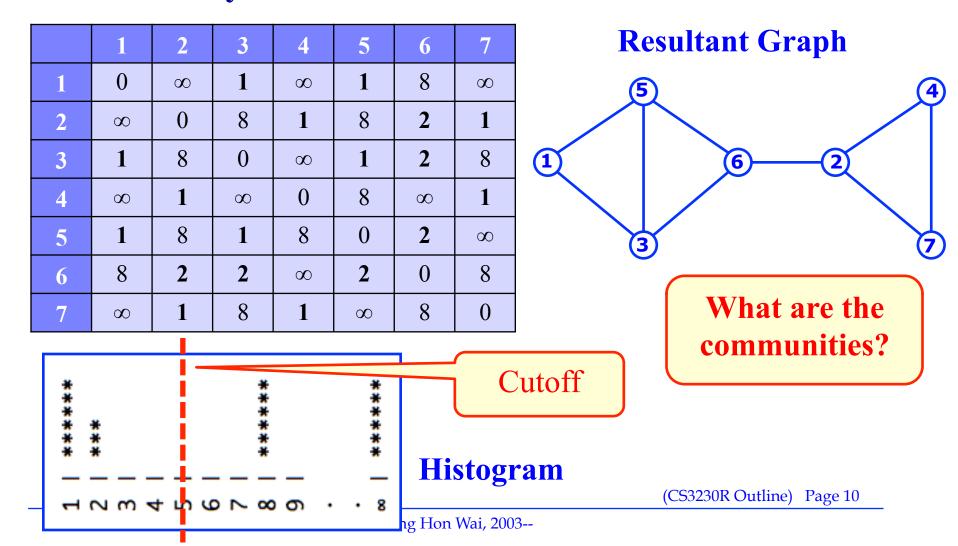
If we set cutoff T=9

Dissimilarity matrix



If we set cutoff T=5 (or 4, or 3)

Dissimilarity matrix

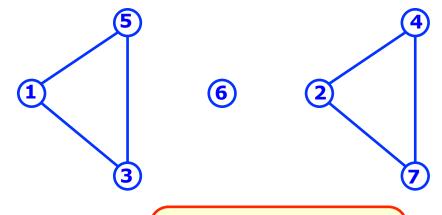


If we set cutoff T=2

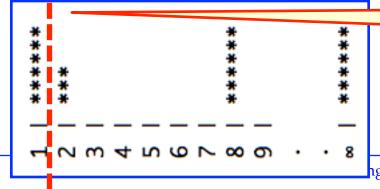
Dissimilarity matrix

	1	2	3	4	5	6	7
1	0	8	1	8	1	8	8
2	8	0	8	1	8	2	1
3	1	8	0	∞	1	2	8
4	8	1	∞	0	8	∞	1
5	1	8	1	8	0	2	8
6	8	2	2	∞	2	0	8
7	∞	1	8	1	∞	8	0

Resultant Graph



What are the communities?



Cutoff

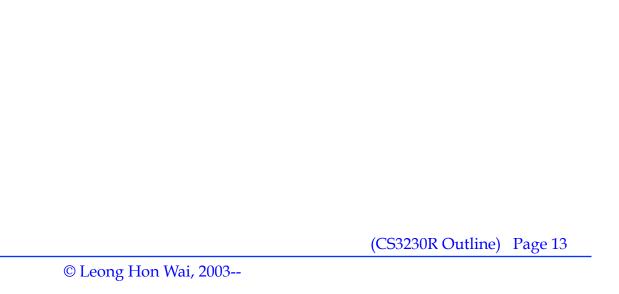
Histogram

(CS3230R Outline) Page 11

ng Hon Wai, 2003--

Conclusion

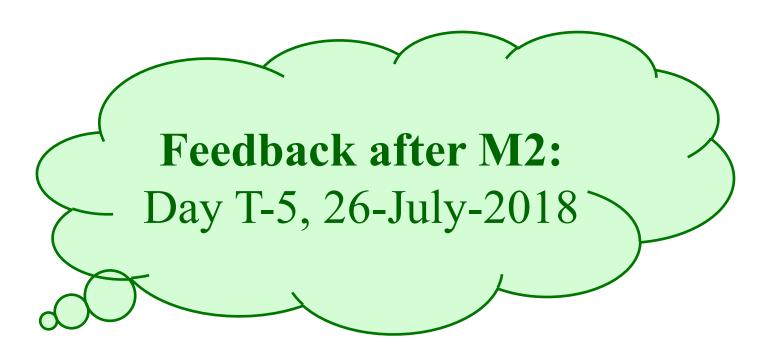
- □ Choosing different cutoffs
 - Changes the graph
 - Changes the communities that you get



CD News (Summer 2018)



School of Computing



Remember...



School of Computing

On 13-July-2018, we said you will do Community Detection algorithm!

How many of you were confident?

(CS3230R Outline) Page 15

☐ Good news --- all groups finished the whole workflow

- ☐ Good news all groups finished the whole workflow
- **□** Presentation
 - * Present the Overall flow,
 - * Don't JUMP straight into results;
- □ Powerpoint should NOT replace poster or the presentation
 - Should augment / give more details

- ☐ Issues with dataset
 - Too many sub-classes
- **□** What is the dissimilarity
 - **Does the numbers make sense**
- ☐ Graph: Undirected vs Directed
 - Found an appropriate algorithm
- ☐ Presentation of the communities
 - how they represent

- **□** What to do with large communities
- ☐ Analysis of your communities
 - Highlight important characteristics
 - * Don't just SAY, show evidence
 - * Those that appear "strange" maybe new insights..

Feedback... (2017)

What if you get some funny result?

- ☐ Like a very large community
 - ***** many smaller communities group together.
- ☐ How to break down large communities?
 - Change Cut-offs
 - * Modify parameters (eg: e, r, Q, cutoff)
 - **Else, modify algorithms ??**
 - **Else, "modify" output** (GN gives output)

Dilemma in Big Data analysis

- □ Don't know what we know
- ☐ Know what we know
- ☐ Know what we don't know
- □ Don't know what we don't know

What can we do?
How can CD help?

Use of Community Detection (CD)

- 1. Big data gives us Large Graphs
- 2. Detect communities in these Graphs
- 3. Analyze the communities detected
 - Develop/confirm relationships between your data, your entities, the groups
 - Derive new insights from community relationships
 - **Learn new things from your Big Data**
 - **Turn into competitive advantage**

New buzzword...



Thank you!