



Large Graph Mining: Power Tools and a Practitioner's guide

Christos Faloutsos

Gary Miller

Charalampos (Babis) Tsourakakis

CMU

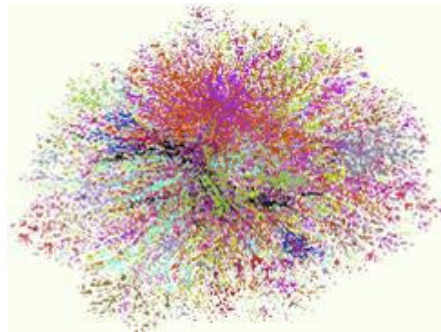


Outline

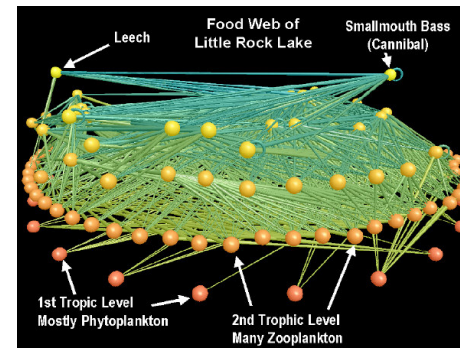
- Introduction – Motivation
- Task 1: Node importance
- Task 2: Community detection
- Task 3: Recommendations
- Task 4: Connection sub-graphs
- Task 5: Mining graphs over time
- Task 6: Virus/influence propagation
- Task 7: Spectral graph theory
- Task 8: Tera/peta graph mining: hadoop
- Observations – patterns of real graphs
- Conclusions



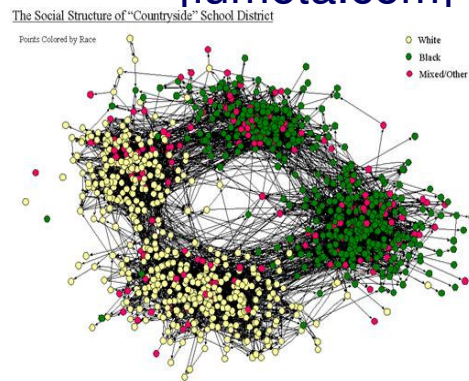
Graphs - why should we care?



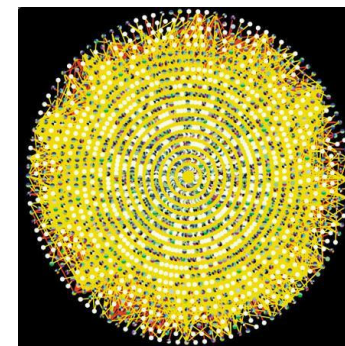
Internet Map
[lumeta.com]



Food Web
[Martinez '91]



Friendship Network
[Moody '01]

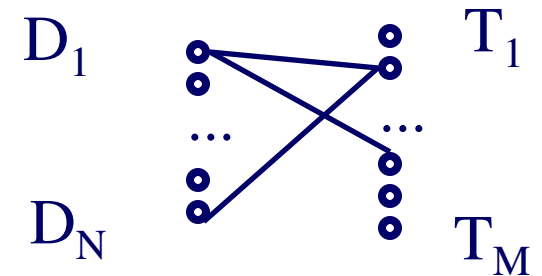


Protein Interactions
[genomebiology.com]



Graphs - why should we care?

- IR: bi-partite graphs (doc-terms)



- web: hyper-text graph
- ... and more:



Graphs - why should we care?

- network of companies & board-of-directors members
- ‘viral’ marketing
- web-log (‘blog’) news propagation
- computer network security: email/IP traffic and anomaly detection
-
- **Any M:N relationship -> Graph**
- **Any subject-verb-object construct: -> Graph**



Graphs and matrices

- Closely related
- Powerful tools from matrix algebra, for graph mining
- (and, conversely, great problems for CS/Math to solve)



Examples of Matrices:

Graph - social network

	John	Peter	Mary	Nick	...
John	0	11	22	55	...
Peter	5	0	6	7	...
Mary
Nick
...



Examples of Matrices:

Market basket

- **market basket** as in Association Rules

	milk	bread	choc.	wine	...
John	13	11	22	55	...
Peter	5	4	6	7	...
Mary
Nick
...



Examples of Matrices:

Documents and terms

	data	mining	classif.	tree	...
Paper#1	13	11	22	55	...
Paper#2	5	4	6	7	...
Paper#3
Paper#4
...



Examples of Matrices: Authors and terms

	data	mining	classif.	tree	...
John	13	11	22	55	...
Peter	5	4	6	7	...
Mary
Nick
...



Examples without networks

- (or, at least, no straightforward network interpretation)



Examples of Matrices: cloud of n-d points

	chol#	blood#	age
John	13	11	22	55	...
Peter	5	4	6	7	...
Mary
Nick
...



Examples of Matrices: sensor-ids and time-ticks

	temp1	temp2	humid.	pressure	...
t1	13	11	22	55	...
t2	5	4	6	7	...
t3
t4
...



Outline

- Introduction – Motivation
- ➔ • Task 1: Node importance
- Task 2: Community detection
- ...
- Conclusions