CS 6824: Hypergraph Algorithms and Applications

T. M. Murali

January 22, 2014

Course Information

- Meet on Mondays and Wednesdays, 4:00pm-5:15pm, Randolph 320.
- Office hours: 9:30am-11:30am, Mondays, and by appointment.
- Course website: `http://courses.cs.vt.edu/~cs6824`. Consult this website regularly. Course schedule is subject to change.
- I will also use Scholar to post some lectures and some papers.

Course Pre-requisites

- Graduate work in any of the following will be useful:
  - Algorithms
  - Machine Learning
  - Data Mining
- Ability to program in one or more of the following languages is important:
  - Python
  - Matlab
  - C++
  - Java

Course Structure

Discuss state-of-the-art research papers.

Course Structure

Discuss state-of-the-art research papers.

- Lectures

Course Structure

Discuss state-of-the-art research papers.

- Lectures
- Student presentations (individual or group)

Course Structure

Discuss state-of-the-art research papers.

- Lectures
- Student presentations (individual or group)
- Invited lectures

Course Structure

Discuss state-of-the-art research papers.

- Lectures
- Student presentations (individual or group)
- Invited lectures
- Class participation

Course Structure

Discuss state-of-the-art research papers.

- Lectures
- Student presentations (individual or group)
- Invited lectures
- Class participation
- Final project: either research project or term paper

Grading

- Presentation: 30%
- Class participation: 30%
- Final project: 40%

Grading

- Presentation: 30%
- Class participation: 30%
- Final project: 40%
- Class participation is very important.

Grading

- Presentation: 30%
- Class participation: 30%
- Final project: 40%
- Class participation is very important.
- Class participation $\neq$ attendance!

Individual Presentations

- Number of papers: each student and I mutually decide a set of 1–2 (and perhaps 3) papers. You can either present one paper in detail (and summarise others) or give equal importance to all papers.
- Time: present for 45 minutes and expect 30 minutes of questions and discussion during the presentation. Be prepared for some discussions to take over your presentation.
- Prepare your presentation well in advance. Practise multiple times.
- Please give me PDF copies of slides (no Microsoft PowerPoint).

Suggestions on Reading and Presenting Papers

- Be sceptical/critical: even papers in the leading journals and conferences have errors or invalid thinking.

Suggestions on Reading and Presenting Papers

- Be sceptical/critical: even papers in the leading journals and conferences have errors or invalid thinking.
- Questions to consider:
  - Are hypergraph representations necessary for this application or computational problem?
  - Is the algorithm good and computationally efficient? Can you improve the technique?
  - Can you mathematically describe the output of the algorithm? Can you provide sketches of the proofs?
  - Don't have to give all details. You can just present the essential ideas.

Suggestions on Reading and Presenting Papers

- Be sceptical/critical: even papers in the leading journals and conferences have errors or invalid thinking.
- Questions to consider:
  - Are hypergraph representations necessary for this application or computational problem?
  - Is the algorithm good and computationally efficient? Can you improve the technique?
  - Can you mathematically describe the output of the algorithm? Can you provide sketches of the proofs?
  - Don't have to give all details. You can just present the essential ideas.
- Read supplementary information. Often has details about the assumptions, the techniques, and the results.

Suggestions on Reading and Presenting Papers

- Be sceptical/critical: even papers in the leading journals and conferences have errors or invalid thinking.
- Questions to consider:
  - Are hypergraph representations necessary for this application or computational problem?
  - Is the algorithm good and computationally efficient? Can you improve the technique?
  - Can you mathematically describe the output of the algorithm? Can you provide sketches of the proofs?
  - Don't have to give all details. You can just present the essential ideas.
- Read supplementary information. Often has details about the assumptions, the techniques, and the results.
- You might have to read other papers in order to present the one assigned to you.

Student Groups For Projects

- Each group has 2–4 members.
- You can form your own groups.
- Try to form groups with students with different backgrounds.

Final Projects

- Research Projects
  - Software + analysis project.
  - We will define a project inspired by the papers you present.
  - I will discuss list of projects within the next few weeks.
  - You can propose a project to me.
  - I will meet each group once a month to monitor progress.
  - You can use Perl, C, C++, Java, Python, R ...

Final Projects

- Research Projects
  - Software + analysis project.
  - We will define a project inspired by the papers you present.
  - I will discuss list of projects within the next few weeks.
  - You can propose a project to me.
  - I will meet each group once a month to monitor progress.
  - You can use Perl, C, C++, Java, Python, R ...
- Hypergraph Library (in Python)
  - You will need to find efficient data structures for hypergraphs.
  - You will implement basic algorithmic problems on hypergraphs.
    - Shortest Paths
    - Random Walks
    - Network Flows
    - Hypergraph Matching
  - This library will be immediately useful for current research projects.

Final Projects

- Research Projects
  - Software + analysis project.
  - We will define a project inspired by the papers you present.
  - I will discuss list of projects within the next few weeks.
  - You can propose a project to me.
  - I will meet each group once a month to monitor progress.
  - You can use Perl, C, C++, Java, Python, R ...
- Hypergraph Library (in Python)
  - You will need to find efficient data structures for hypergraphs.
  - You will implement basic algorithmic problems on hypergraphs.
    - Shortest Paths
    - Random Walks
    - Network Flows
    - Hypergraph Matching
  - This library will be immediately useful for current research projects.
- The software has to run on Linux!

Final Term Paper

- Propose hypergraphs for a new application/area.
- Discuss a superset of the papers you present in class or a group of other papers.
- You can propose additional papers.

Final Term Paper

- Propose hypergraphs for a new application/area.
- Discuss a superset of the papers you present in class or a group of other papers.
- You can propose additional papers.
- The electronic document you submit must be in PDF format. Please do not give me a Microsoft Word document.

Sources of Information

- **There is no textbook for the course.**
- Related book: *Hypergraphs, Volume 45: Combinatorics of Finite Sets (North-Holland Mathematical Library)*, Berge, 1989.

Sources of Information

- **There is no textbook for the course.**
- Related book: *Hypergraphs, Volume 45: Combinatorics of Finite Sets (North-Holland Mathematical Library)*, Berge, 1989.
- Our main sources: conferences and journals.
- `http://www.citeulike.org/user/tmmurali/tag/2014-spring-cs6824-hypergraphs`

Networks

Networks appear in numerous applications.

Networks

Networks appear in numerous applications.

- Biological Networks
- Social Networks
- Virtual Enterprise Networks
- Transportation Networks

Networks are typically represented as a graph $G = (V, E)$ with a set V of nodes and pairwise connections between nodes (E).

Networks

Networks appear in numerous applications.

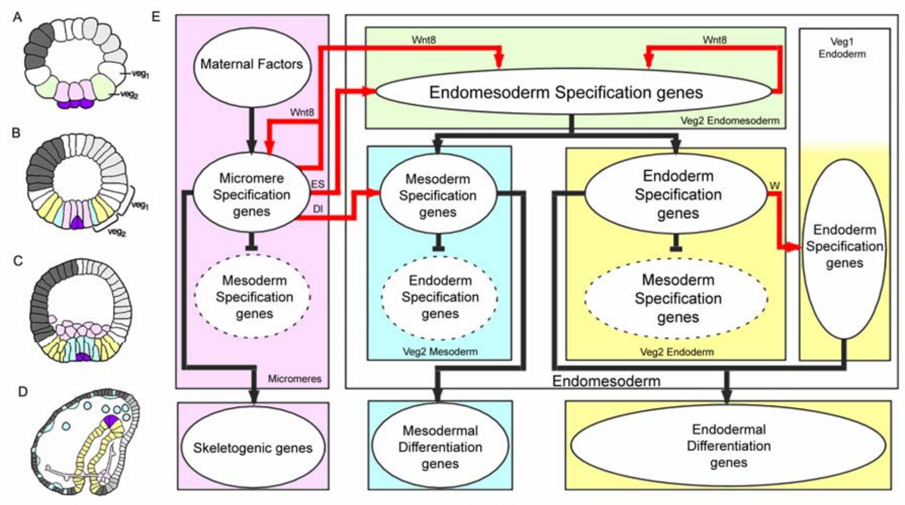
- Biological Networks
- Social Networks
- Virtual Enterprise Networks
- Transportation Networks

Networks are typically represented as a graph $G = (V, E)$ with a set V of nodes and pairwise connections between nodes (E).

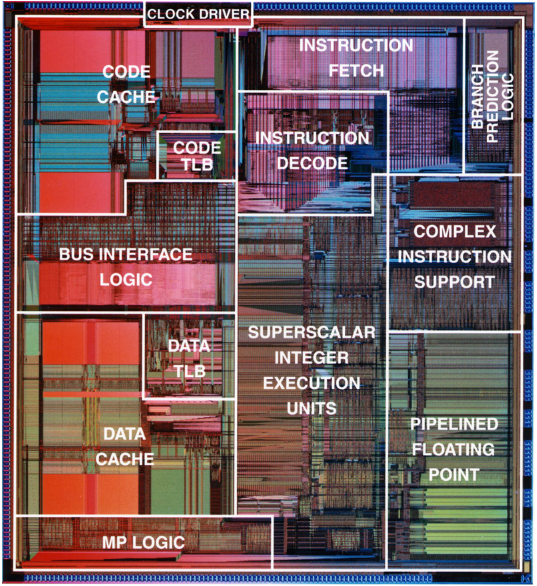
Are graph representations of networks enough?

A Cell

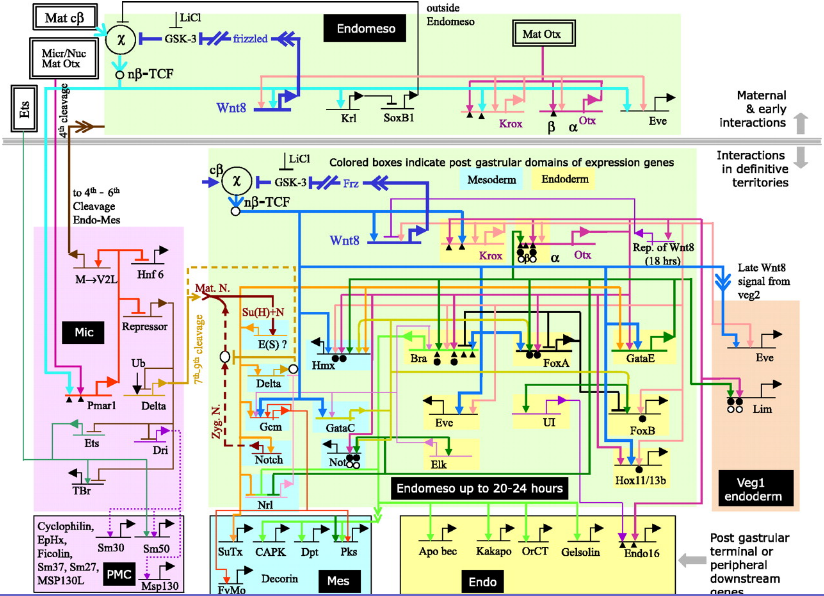
A Cell is a Modular



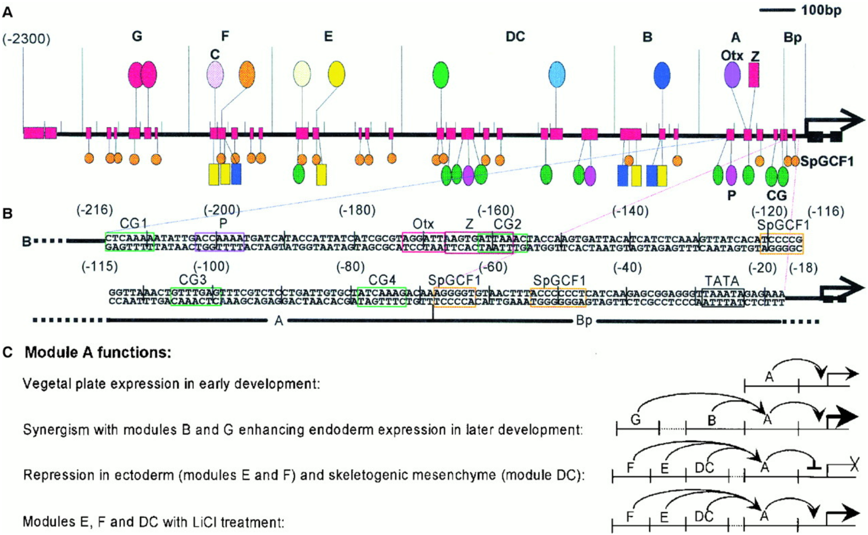
A Cell is a Modular



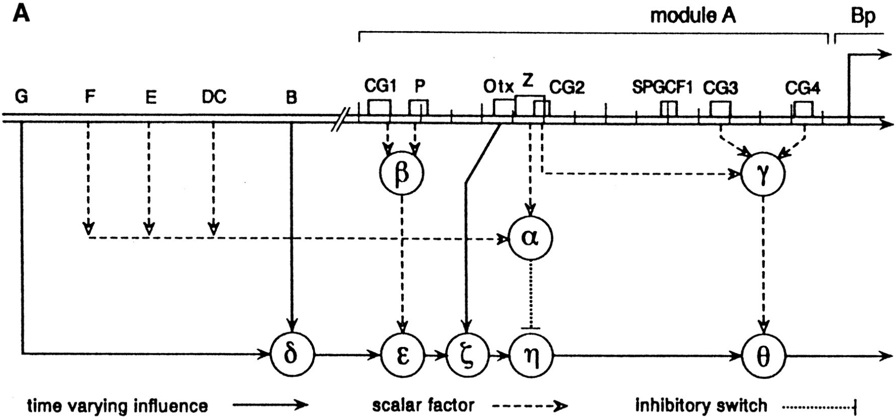
A Cell is a Modular Network



A Cell is a Modular Network



A Cell is a Modular Network that Computes

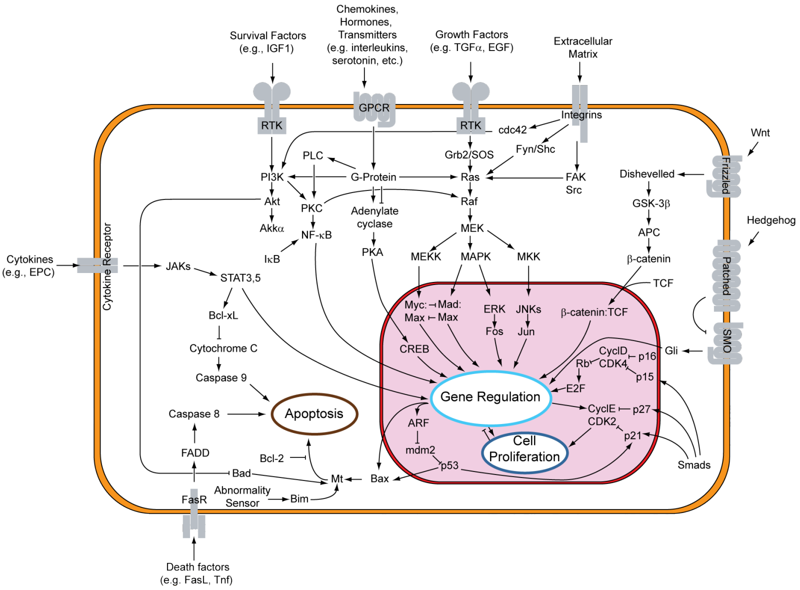


A Cell is a Modular Network that Computes

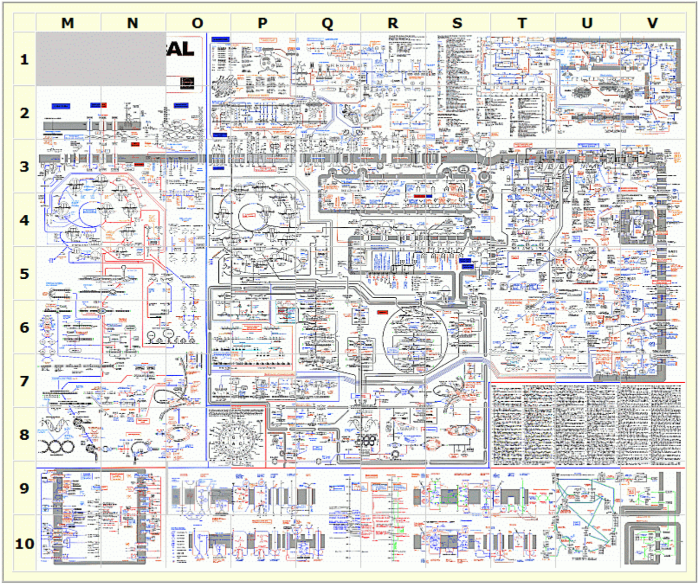
B

if (F = 1 or E = 1 or CD = 1) and (Z = 1) $\alpha = 1$ **else** $\alpha = 0$	Repression functions of modules F, E, and DC mediated by Z site
if (P = 1 and CG_1 = 1) $\beta = 2$ **else** $\beta = 0$	Both P and CG_1 needed for synergistic link with module B
if (CG_2 = 1 and CG_3 = 1 and CG_4 = 1) $\gamma = 2$ **else** $\gamma = 1$	Final step up of system output
$\delta(t) = B(t) + G(t)$	Positive input from modules B and G
$\varepsilon(t) = \beta * \delta(t)$	Synergistic amplification of module B output by CG_1-P subsystem
if ($\varepsilon(t) = 0$) $\xi(t) = Otx(t)$ **else** $\xi(t) = \varepsilon(t)$	Switch determining whether Otx site in module A, or upstream modules (i.e., mainly module B), will control level of activity
if ($\alpha = 1$) $\eta(t) = 0$ **else** $\eta(t) = \xi(t)$	Repression function inoperative in endoderm but blocks activity elsewhere
$\Theta(t) = \gamma * \eta(t)$	Final output communicated to BTA

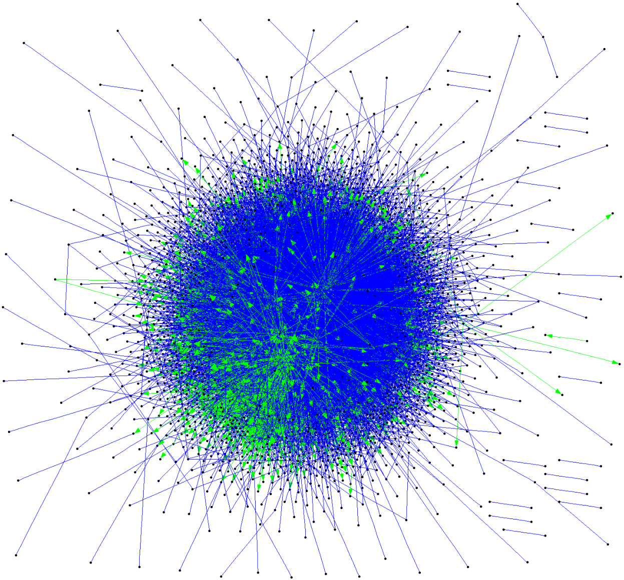
Biological Networks are Complex



Biological Networks are Complex

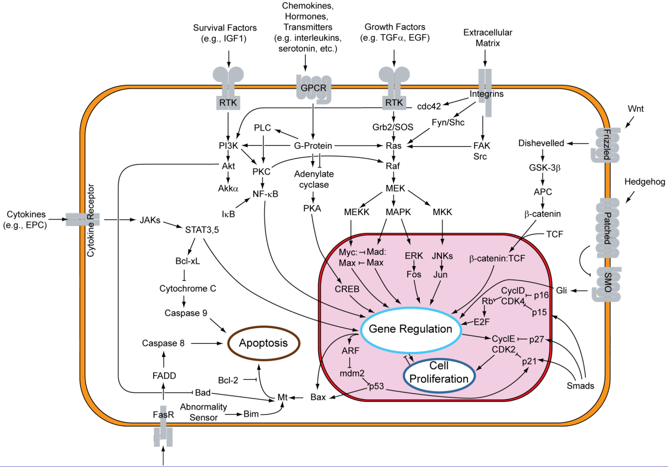


Biological Networks are Complex



Biological Networks as Graphs

- Protein A physically interacts with protein B
- Protein A activates protein B
- Protein A inhibits protein B



Biological Networks as Graphs

- Protein A physically interacts with protein B
- Protein A activates protein B
- Protein A inhibits protein B

But what about:

- **Complexes**
  - The Complex A/B activates protein C
  - Protein A is required to separate the components of complex B/C

Biological Networks as Graphs

- Protein A physically interacts with protein B
- Protein A activates protein B
- Protein A inhibits protein B

But what about:

- **Complexes**
  - The Complex A/B activates protein C
  - Protein A is required to separate the components of complex B/C
- **Combinatorial Regulation**
  - Proteins A and B together activate protein C
  - Protein A activates protein B, but only if protein C is not present

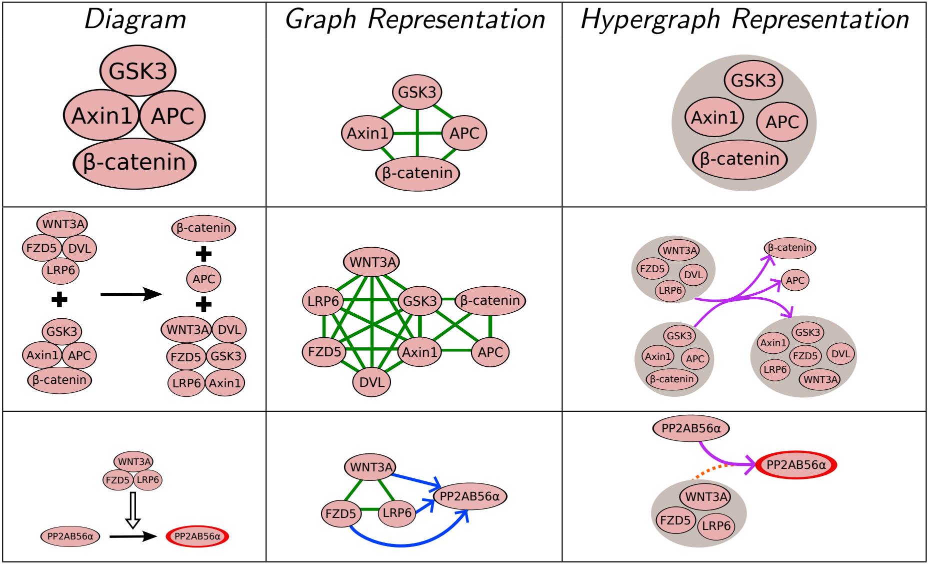


Diagram
Graph Representation
Hypergraph Representation
GSK3
Axin1
APC
β-catenin
WNT3A
FZD5
DVL
LRP6
PP2AB56α

Social Networks

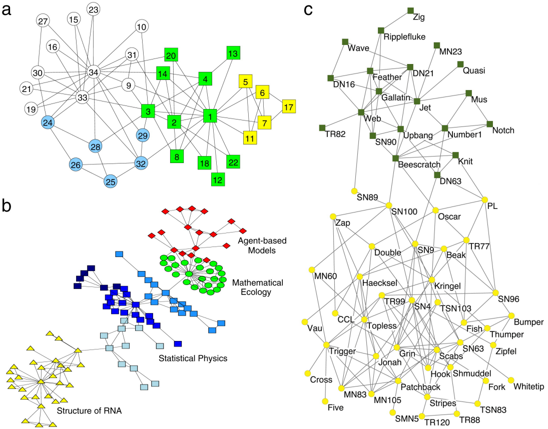
- *Social Networks* are graphs whose nodes are individuals and edges are pairwise relationships between individuals.
  - Websites (FaceBook, Twitter, LinkedIn, Digg, Flickr, ...)
  - Authorship networks
  - Email and instant messaging networks

Social Networks

- *Social Networks* are graphs whose nodes are individuals and edges are pairwise relationships between individuals.
  - Websites (FaceBook, Twitter, LinkedIn, Digg, Flickr, ...)
  - Authorship networks
  - Email and instant messaging networks
- *Affiliation Networks* are graphs with two types of nodes – individuals and societies/groups – with edges between them.
  - Forms a bipartite graph.
  - Can recover social network by "folding" the graph.

Community Detection in Social Networks

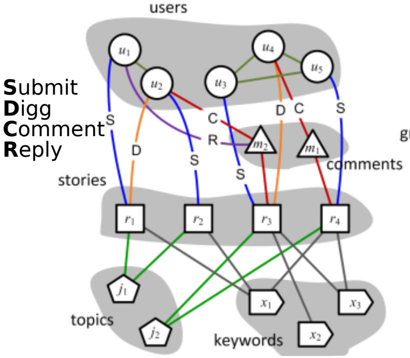
Fig. 2. from `http://www.sciencedirect.com/science/article/pii/S0370157309002841` Community structure in social networks (colors attempt to find communities).



- (a) Zacharys karate club, a standard benchmark in community detection.
- (b) Collaboration network between scientists working at the Santa Fe Institute. The colors correspond quite closely to research divisions of the institute. Further subdivisions correspond to smaller research groups, revolving around project leaders.
- (c) Lusseaus network of bottlenose dolphins. The partition matches the biological classification of the dolphins proposed by Lusseau.

Social Networks are Complex

Primary Actions and Related Media Objects in Digg



Primary Actions and Related Media Objects in Flickr

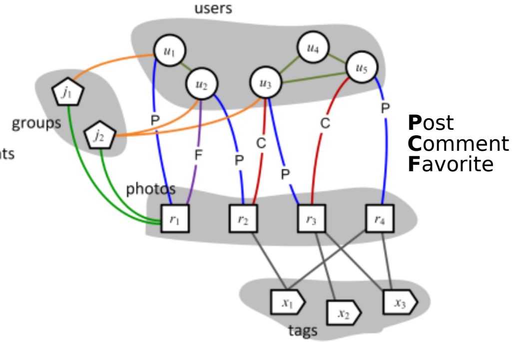


Figure from `http://www.citeulike.org/group/3240/article/5663323`.

Vehicle Rotation Planning

In train scheduling, the vehicle rotation planning problem finds an assignment of each trip in a timetable to a follow-on trip which will be serviced by the same vehicle.

`http://www.citeulike.org/group/3240/article/12926045`

Bicluster Visualization

"Given a set of discovered submatrices of interests, how can we order the rows and columns of the data matrix to best display these submatrices and their relationships?"

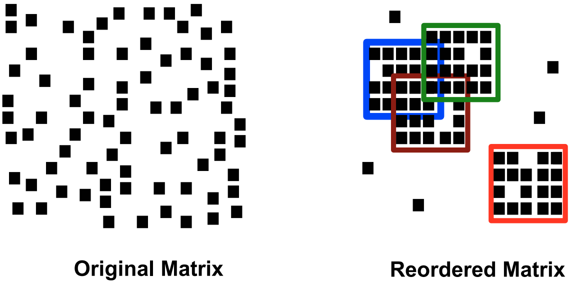


Figure 1. An example of matrix pattern visualization

Image from
`http://www.citeulike.org/group/3240/article/12926044`

Homework

- Each student must find two new papers on hypergraphs: one algorithmic and one application. new $\equiv$ not on CiteULike
- Send me URLs to both papers.
- Prepare a 5 minute presentation on either paper for the class on Monday, January 27, 2014.
  - Why did you find this paper interesting and/or relevant?
  - Algorithmic: What problem does the paper solve?
  - Application: Where do hypergraphs play a role?
- Send me the PDF slides by noon on January 27, 2014.
- I will compile the slides into one presentation to streamline the class.