open.michigan

Unless otherwise noted, the content of this course material is licensed under a Creative Commons Attribution 3.0 License.
http://creativecommons.org/licenses/by/3.0/

Copyright 2008, Lada Adamic

You assume all responsibility for use and potential liability associated with any use of the material. Material contains copyrighted content, used in accordance with U.S. law. Copyright holders of content included in this material should contact open.michigan@umich.edu with any questions, corrections, or clarifications regarding the use of content. The Regents of the University of Michigan do not license the use of third party content posted to this site unless such a license is specifically granted in connection with particular content objects. Users of content are responsible for their compliance with applicable law. Mention of specific products in this recording solely represents the opinion of the speaker and does not represent an endorsement by the University of Michigan. For more information about how to cite these materials visit http://michigan.educommons.net/about/terms-of-use.

University of Michigan

CC BY

School of Information
University of Michigan

Introductory social network analysis with Pajek

Lada Adamic

Outline

- In Pajek
  - visualization and layouts
  - concepts of centrality: degree,betweenness, closeness
  - connected components
  - snowball sampling
  - one mode projections of bipartite graphs
  - thresholding weighted graphs
- Fetching you facebook network
  - What is an egonetwork?

Using Pajek for exploratory social network analysis

- Pajek – (pronounced in Slovenian as Pah-yek) means ‘spider’

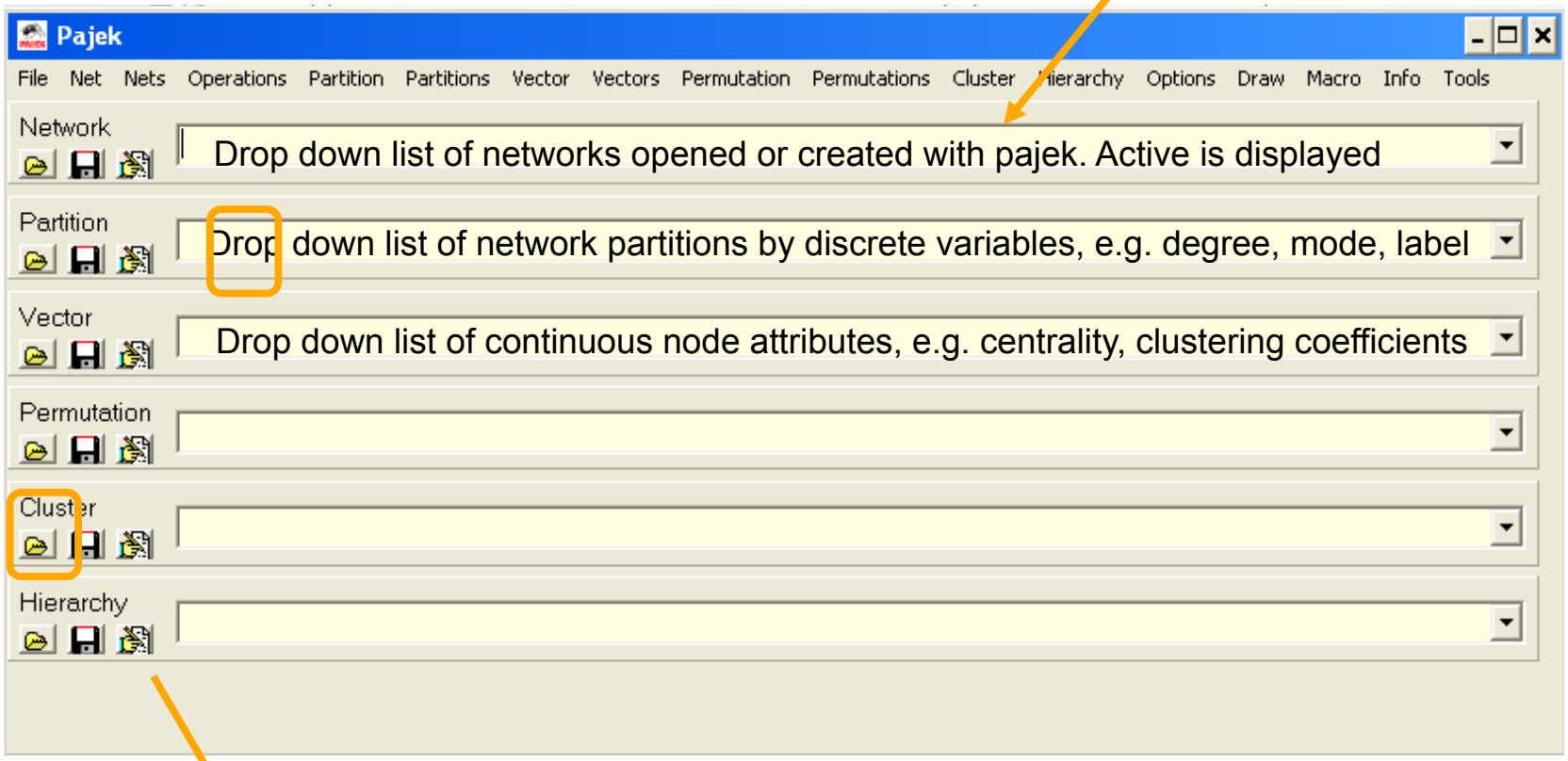
- website: http://vlado.fmf.uni-lj.si/pub/networks/**pajek**/
- wiki: http://pajek.imfm.si/doku.php
  - download application (free)
  - tutorials
  - lectures
  - data sets

- Windows only (works on Linux via Wine, Mac via Darwine)

- helpful book: ‘Exploratory Social Network Analysis with Pajek’ by Wouter de Nooy, Andrej Mrvar and Vladimir Batagelj

Pajek: interface

we'll use today

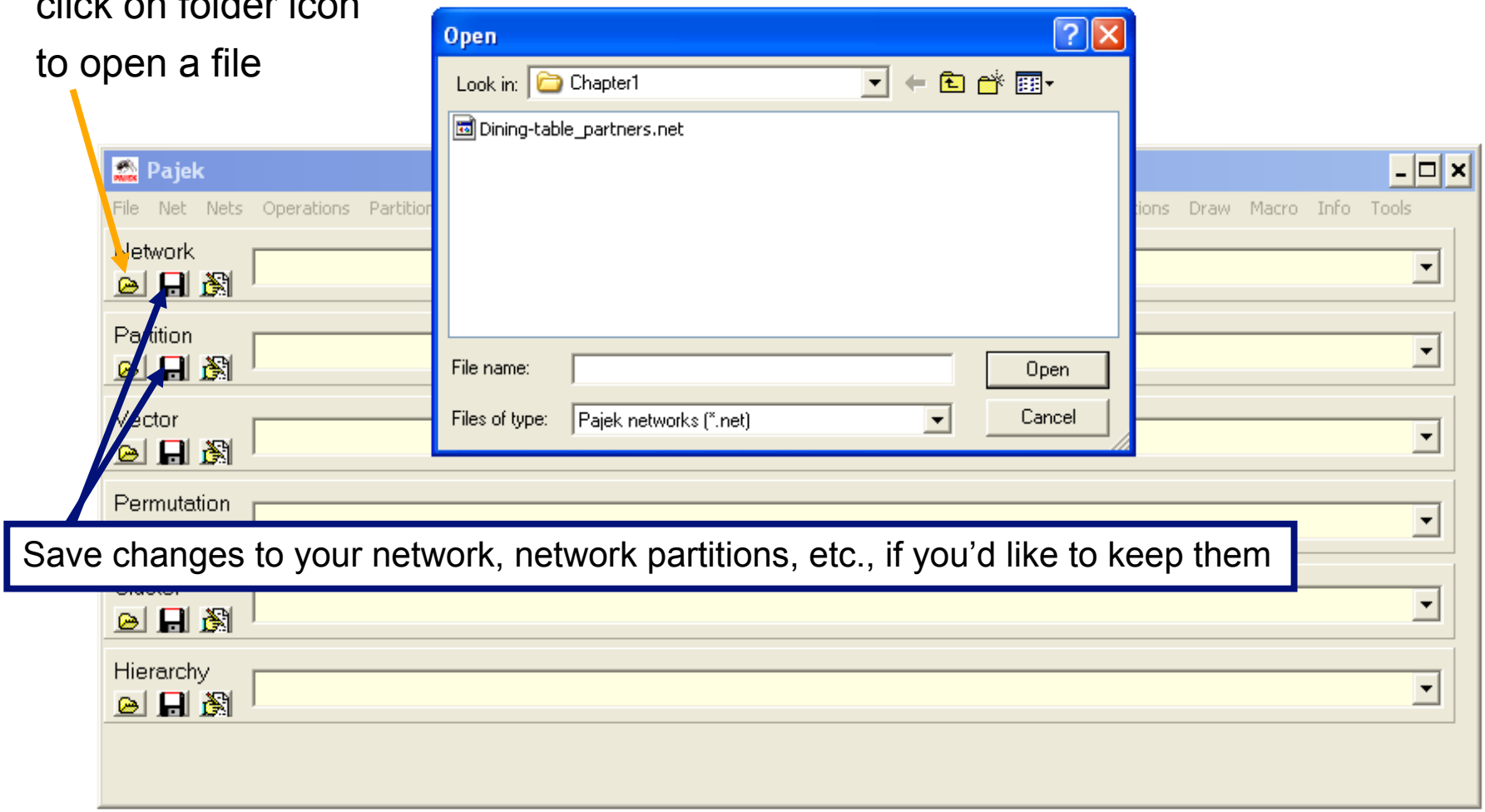


can be used for clustering

Source: Pajek, Free for noncommercial use - http://pajek.imfm.si/doku.php?id=download

Pajek: opening a network file

click on folder icon
to open a file

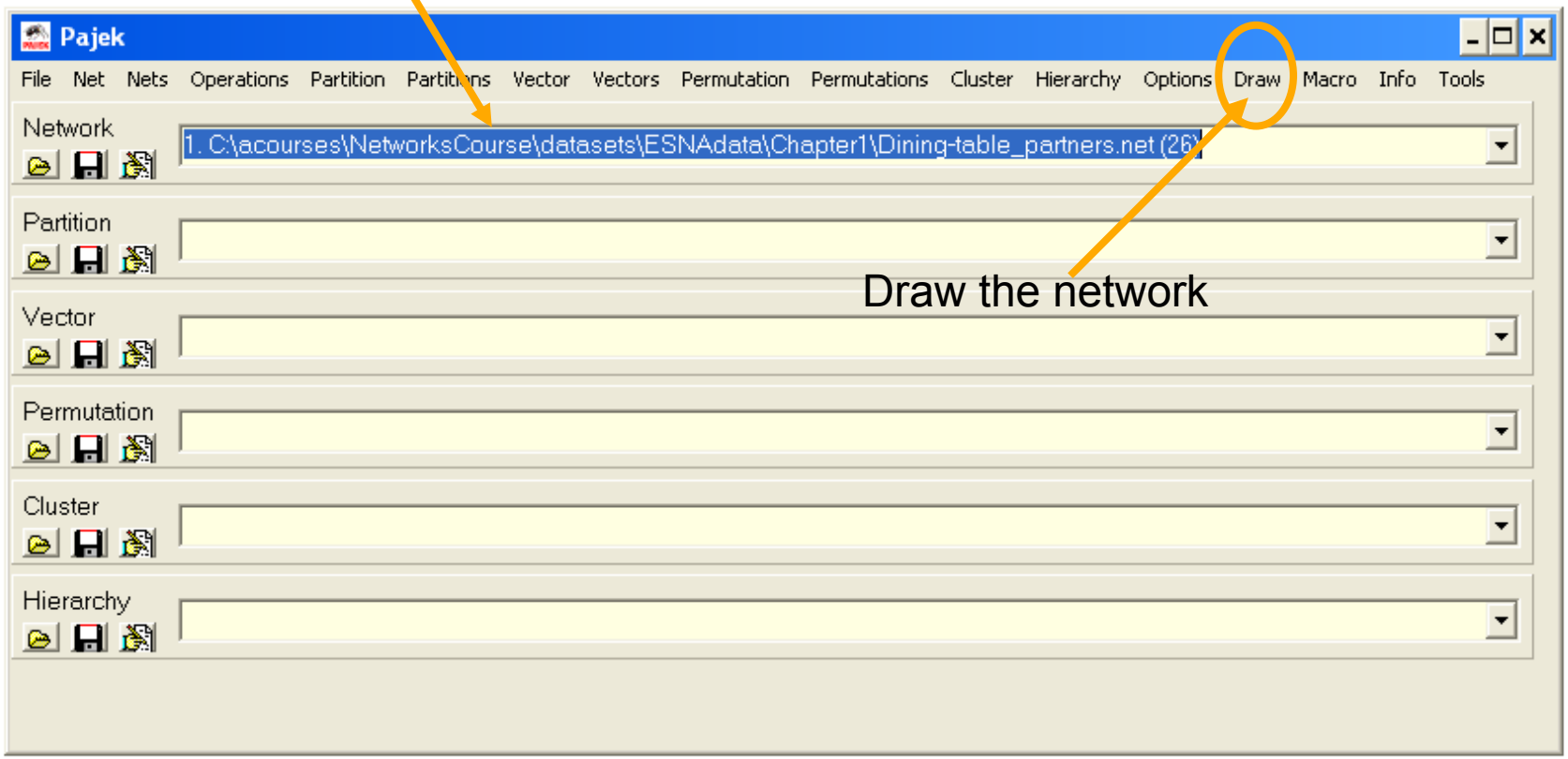


Save changes to your network, network partitions, etc., if you'd like to keep them

Source: Pajek, Free for noncommercial use - http://pajek.imfm.si/doku.php?id=download

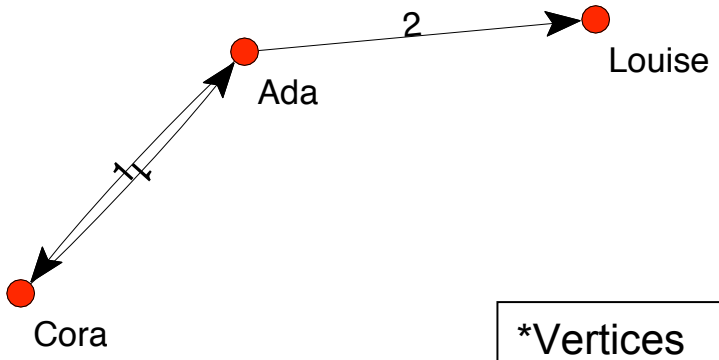
Pajek: working with network files

- The active network, partition, etc is shown on top of the drop down list



Source: Pajek, Free for noncommercial use - http://pajek.imfm.si/doku.php?id=download

Pajek data format



number of vertices

vertex x,y,z coordinates (optional)

```
*Vertices     26
     1 "Ada"          0.1646   0.2144   0.5000
     2 "Cora"         0.0481   0.3869   0.5000
     3 "Louise"       0.3472   0.1913   0.5000
     ..

*Arcs
     1 3 2 c Black
     ..
*Edges
     1 2 1 c Black
     ..
```

directed edges

from Ada(1) to Louise(3) as choice “2” and color Black

undirected edges

between Ada(1) to Cora(2) as choice “1” and color Black

Pajek: Let's get started

- Opening a network
  - File > Network > Read
- Visualization
  - Draw > Draw
- Essential measurements
  - …

Pajek: opening a file

- A planar graph and layouts in Pajek
- Download the file 'planarnet.net' from the ctools website.

- Open it in Pajek by either clicking on the yellow folder icon under the word "Network" or by selecting File>Network>Read from the main menu panel
- A report window should pop up confirming that the graph has been read and the filename and location will be displayed in the 'active' position of the network dropdown list

Pajek: visualization & manual positioning

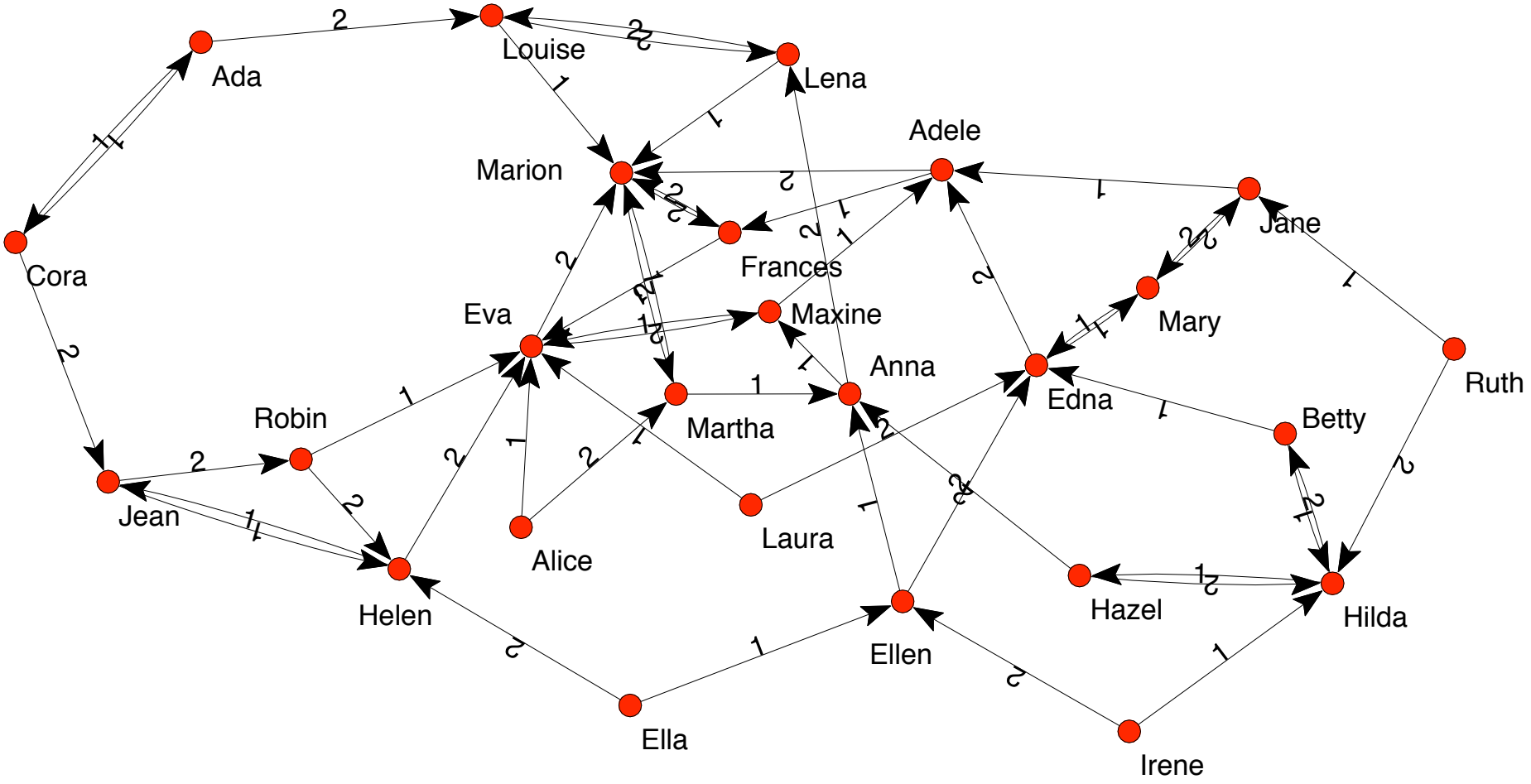
- Visualize the network using Pajek's Draw>Draw command from the main menu panel.
- This will bring up the 'draw' window with its own menu bar at the top
- Reposition the vertices by clicking on them and holding down the mouse button while dragging them to a new location. Continue doing this until you have shown that the graph is planar (no edges cross have to cross)
- (If you think this is really fun to do in your spare time, go to http://www.planarity.net)

Pajek: visualization & layout algorithms

- Now let Pajek do the work for you by selecting from the draw toolbar several layout algorithms under 'Layout>Energy'.
- Why did you select the layout algorithm you did?
- Did the layout leave any lines crossed? If you were to do this assignment over, what order would you do it in?

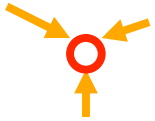
a directed network

- girls' school dormitory dining-table partners (Moreno, *The sociometry reader*, 1960)
- first and second choices shown



Node centrality (importance)

- Node network properties
  - from immediate connections
    - indegree
      how many directed edges (arcs) are incident on a node

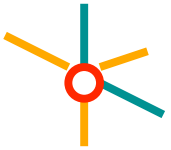


indegree=3

    - outdegree
      how many directed edges (arcs) originate at a node

outdegree=2

    - degree (in or out)
      number of edges incident on a node



degree=5

  - from the entire graph
    - centrality (betweenness, closeness)

Centrality: degree

- More on degree and other centrality measures in the next lecture…
- Degree: calculate it
  - Net > Partitions > Degree
- Visualize degree centrality
  - Draw Draw > Vector
  - If nodes are not the right size, use resize option
    - Draw Options > Size of Vertices

Centrality: betweenness

- The fraction of all directed paths between any two vertices that pass through a node

betweenness of vertex i

paths between j and k that pass through i

$$C_B(n_i) = \sum_{j,k} g_{jk}(i) / g_{jk}$$

all paths between j and k

- Normalization
  - undirected: (N-1)*(N-2)/2
  - directed graph: (N-1)*(N-2) e.g.

$$C_B^{'}(n_i) = C_B(i) / [(N-1)(N-2)]$$

- Pajek: Net > Vector > Centrality > Betweenness

Centrality: closeness

- How close the vertex is to others
  - depends on inverse distance to other vertices

$$C_c(i) = \left[\sum_{j=1}^{g} d(i\ ,j) \right]^{-1}$$

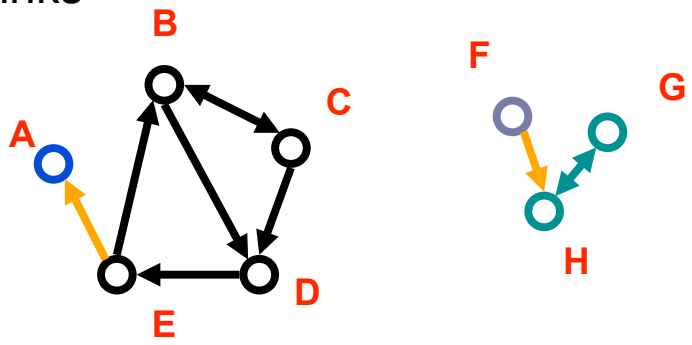
- Normalization

$$C_C^{'}(i) = (C_C(i))(n-1)$$

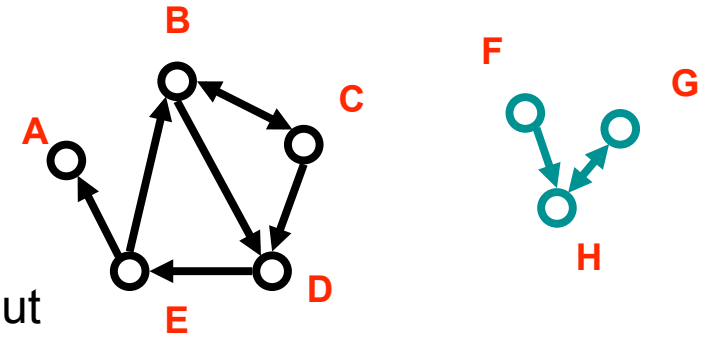
- Pajek: Net > Vector > Closeness …

Network metrics: connected components

- Strongly connected components
  - Each node within the component can be reached from every other node in the component by following directed links
    - Strongly connected components
      - B C D E
      - A
      - G H
      - F



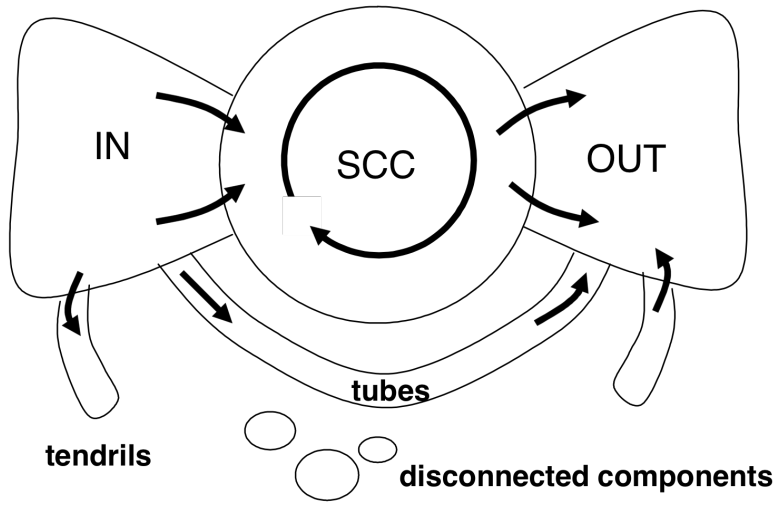
- Weakly connected components: every node can be reached from every other node by following links in either direction
  - Weakly connected components
    - A B C D E
    - G H F



- In undirected networks one talks simply about ‘connected components’

The bowtie model of the Web
Broder et al. (1999)

- SCC (strongly connected component):
  - can reach all nodes from any other by following *directed* edges
- IN
  - can reach SCC from any node in 'IN' component by following directed edges
- OUT
  - can reach any node in 'OUT' component from SCC
- Tendrils and tubes
  - connect to IN and/or OUT components but not SCC
- Disconnected
  - isolated components



Let's try this on the dining table partners

- Suppose each girl will share her dish (and any dish that is shared with her), with only her 1st and 2nd choice

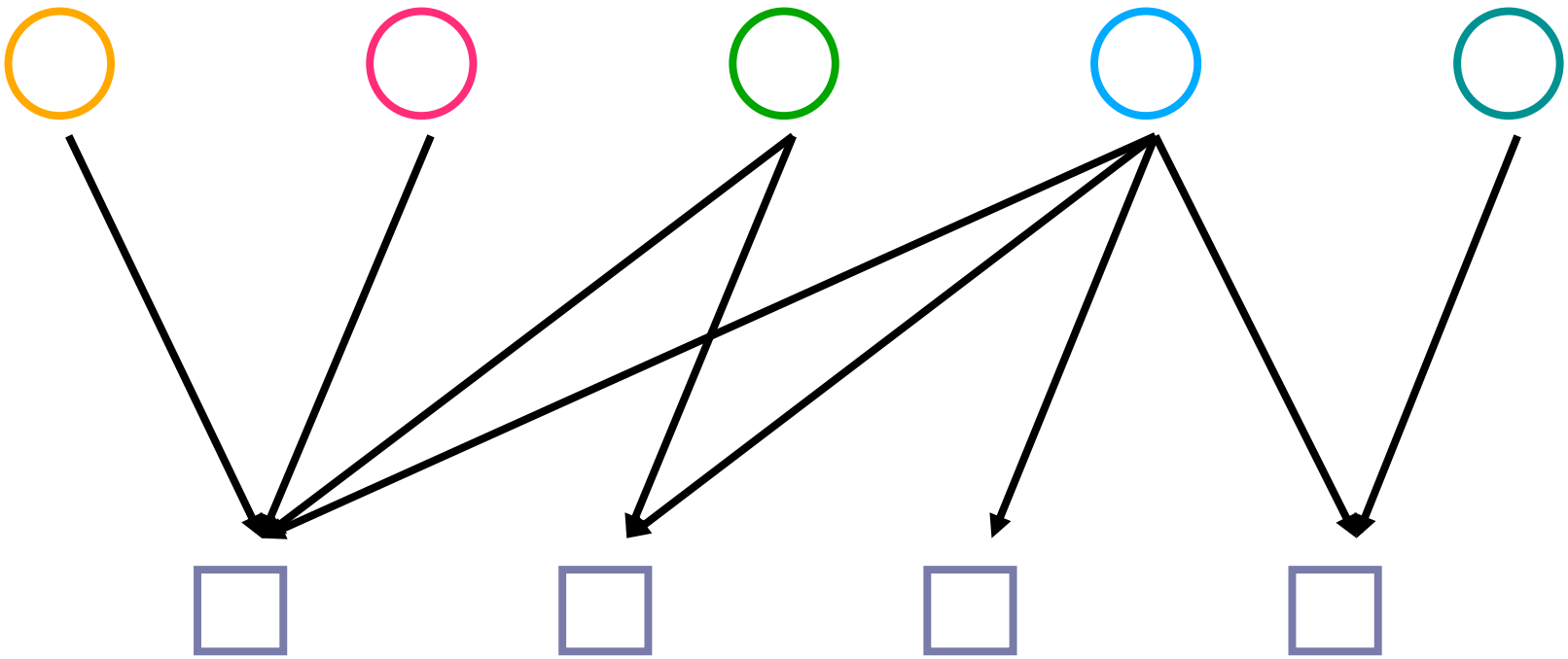
  - Find groups of girls who can all sample each other's dishes
    - Net>Components>Strong and Draw>Draw-partition
  - Which girls will get to sample no other dishes but their own?
    - Net>Partitions>Degree>Input
  - Display the network of strongly connected components
    - Operations>Shrink Network>Partition and Draw>Draw-partition
  - Identify the bowtie structure
    - Net>Partitions>Bow-Tie
      - What part of the bowtie is missing? What does this imply for the circulation of the different dishes?

Snowball Sampling

- **Dining partners:**
  You are a prince who just met an enchanting young lady at a ball, but she left at the stroke of midnight and left a shoe behind. Now you'd like to find the shoe's owner. All you know about her is that she lives in this particular girls' dorm. The headmistress won't let you talk to the girls, so the only way you can find your princess is to covertly ask the one girl you know, Ella, to introduce you to her two favorite friends. Once you know her friends, you can ask them to introduce you to their two favorite friends, etc. This is the snowball sampling technique we covered in class.

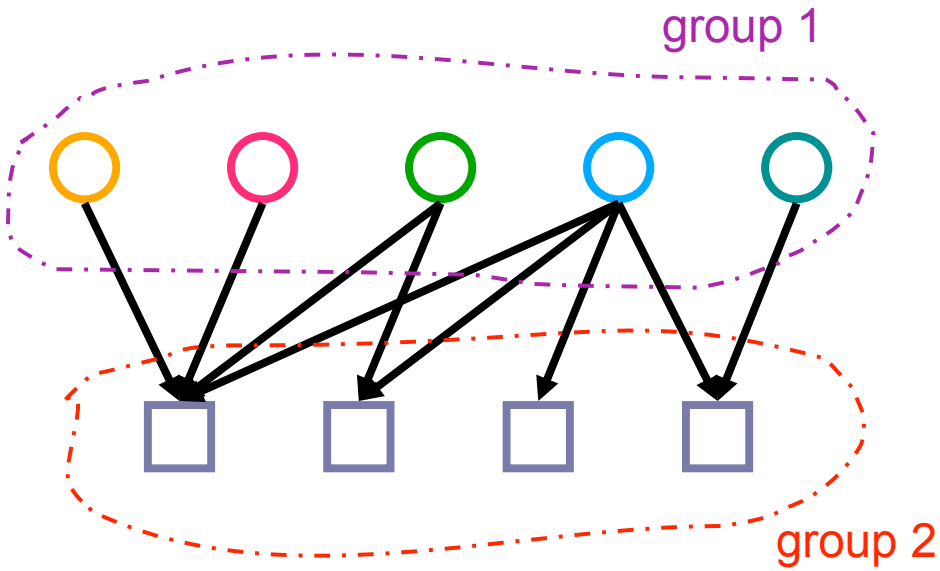
  Highlight the vertices that you will reach using snowball sampling
  - Net > K-Neighbors > ...
    - Which girls will you not find using snowball sampling starting with Ella?

Bipartite networks

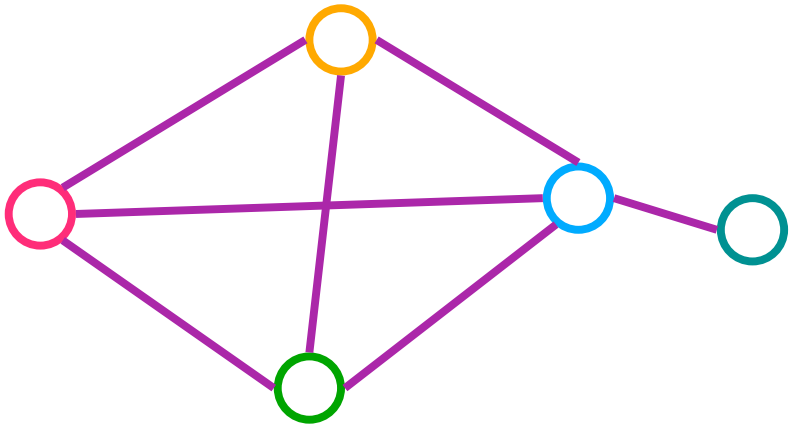


going from a bipartite to a one-mode graph

- Two-mode network



- One mode projection
  - two nodes from the first group are connected if they link to the same node in the second group
  - some loss of information
  - naturally high occurrence of cliques



Let's try it on the actors network

- Open actorsandmovies.net
- Create a 2-mode partition (newer versions of Pajek do this automatically)
  - Net>Partition>2-Mode
- Draw the network
  - Draw>Draw-Partition (the two classes of nodes should be colored differently)
  - If labels are not shown, add them: Options>Mark Vertices Using>Labels
- Experiment with different layout algorithms
  - "Layout" menu in the draw window

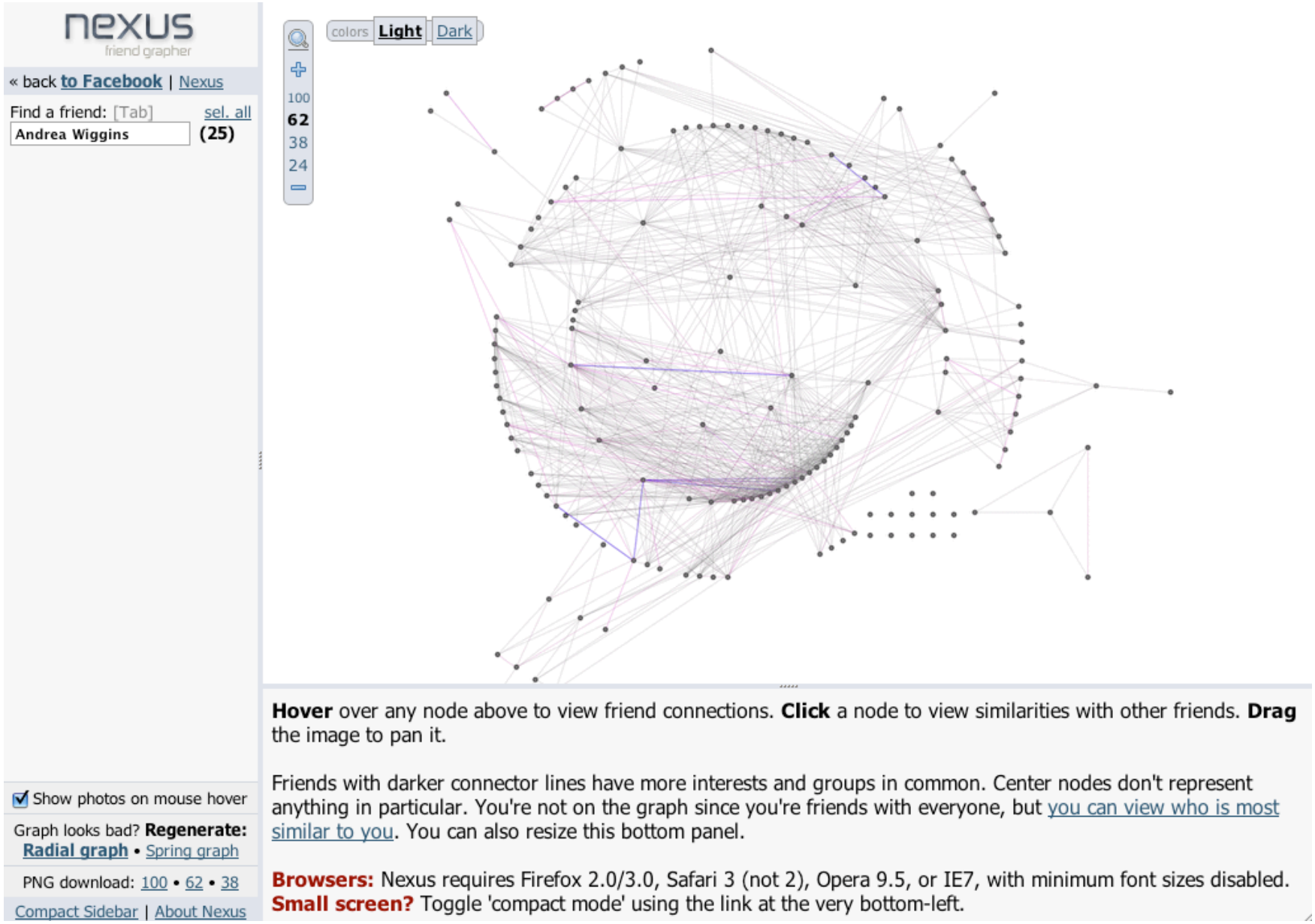
actors network (continued)

- Transform the network into a one-mode network
  - Net>Transform>2-Mode to 1-Mode>Rows
- Draw the network
  - Draw>Draw
    - Qualitatively compare the structure of the 2-Mode to the 1-Mode network. Is there a loss of information?
- Show the weights on each edge
  - Options>Lines>Mark Lines>with Values
    - What do the values represent?
- Compute the unweighted degree of each node
  - Net>Partitions>Degree>All
- Draw the network using
  - Draw>Draw-vector
    - How is the degree represented?
- Add the vector value to each vertex (it will be the degree/(max possible degree))
  - Options>Mark Vertices Using>Vector Values
    - Who are the most important actors using this measure?

actors network (continued)

- How does the boundary of the network (i.e. who is included) affect who is found to be most central?
  - Load the file actorsandmoviesWithGere.net. It contains one extra actor, Richard Gere. Repeat the above procedure.
  - In the 1-mode network of actors, is there a change in who is most central?
  - What does this tell you about biases and boundaries in sample selection?
- Remove all edges between actors who have costarred in fewer than 3 movies
  - Net>Transform>Remove>Lines with value>Lower than
    - Which actors comprise the central core of this network

Grabbing your data from Facebook via Nexus



Pajek: wrap up

- Used frequently by sociologists
  - UCInet is comparable and arguably more user friendly
- Extensive functionality
  - But not extendable
- What we covered
  - visualization
  - node properties: centrality
  - connected components
  - k-neighbors
  - converting two-mode networks to one-mode
  - thresholding the network