# CS 780 Data Mining for Multimedia Data 

## Social Network Graph Mining

Dr. Jessica Lin

## Betweenness

$\square$ Betweenness of an edge $(a, b)$ : the number of pairs of nodes $x$ and $y$ such that the edge $(a, b)$ lies on the shortest path between $x$ and $y$.

- If there are more than one shortest path, the edge $(a, b)$ is credited with the fraction of those shortest paths that include the edge (a, b)
- Need to calculate the number of shortest paths going through each edge
- Girvan-Newman Algorithm


## Girvan-Newman (GN) Algorithm

- Step 1: Convert graph to the BFS presentation starting at node X.
- Label each node by the number of shortest paths that reach it from the root.
$\star$ Start by labeling the root 1
* For each level, label each node $Y$ by the sum of the labels of its parents
- For each edge e, calculate the sum over all nodes $Y$ of the fraction of shortest paths from the root $X$ to $Y$ that go through e
$\star$ Each leaf in the DAG gets a credit of 1
$\star$ Each node that is not a leaf gets a credit equal to $\{1+$ the sum of the credits of the DAG edges from that node to the level below\}
$\star$ Each edge e gets credit from its child node. If the child node has $n$ parent edges, then each edge gets $1 / n$ credit


## Example


(1)

Level 1

Level 2

Level 3
(2)

(3)


## Girvan-Newman Algorithm

■ Intuitively, why should this work? Analogy:

* Network of N nodes: nodes are towns, edges are roads
* Place N-1 cars on each node; each one to a town
* Each road gets a point when a car drives on it
* Remove the highest ranked road - interstate highway
$\star$ Repeat the process
* First we'll remove all interstates (leaving state roads)
$\star$ Then state roads will be removed, leaving county roads, then suburban roads, etc
$\star$ After we each set of levels, we get a more fine-grained division of communities


## Other Slides

- Challenges in Mining Large-Scale Social Network Data: http://cs.stanford.edu/people/jure/talks/networks-icdm-dec12.pdf * (more here: http://cs.stanford.edu/people/jure/talks/)

■ Discovering Clusters in Networks:
http://snap.stanford.edu/class/cs246-2012/slides/11-graphs.pdf

