16-06-2018: Social network analysis

- Huy Le Hoy Phuoy -
HUS, VNU
Chief scientist of FPT

Network → module → cluster → community

SNAP - standford
social network analysis platform

Homophily same properties → connect

Triadic closure:

Clustering coefficient

Power law degree distribution:
$P(k) \sim k^{-\gamma}$ $2 \leq \gamma \leq 3$

Measure of centrality:
prestige
closeness centrality
proximity prestige
betweenness centrality for node
edge

COMMUNITY DETECTION

- spectral clustering *
  - structural properties of real network
  - "spectral representation"
    embed nodes into a multidimensional space R^d

node i → $\vec{v}$ d-dimensions
R^d

eigen vector
value

"spectrum" of matrix

matrix representation (adjencency matrix) A
degree matrix D
laplacian matrix L = D - A

x :

d - 1

cinnamon

nodes
. . .
node n → $Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ R^n

Edge weight $W = (w_{ij})_{n \times n}$

minimize
Tim theo so (y) $J(y) = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (y_i - y_j)^2$

$\lambda_{ii} = \sum_{j=1}^{n} w_{ij}$ sum of weight

diagonal matrix $\Lambda \begin{pmatrix} \lambda_{11} & \\ & \lambda_{nn} \end{pmatrix}$

$L = \Lambda - W$

rewrite → $J(y) = 2y^T L y$ } → solve optimize problem

normalize spectral clustering
$y^T \Lambda y = 1$

$\Lambda^{-1} L y = \lambda y$ → eigenvalue, → eigenvector

Kernighan Lin alg

heuristic
NP - hard
local optimize

cinnamon

init partition graph
↓
improve
1. exchange node between partition
2 - internal cost I_i
3 - External cost E_i

moving (node i) gain $D_i = E_i - I_i$
exchange gain $J_{ij} = D_i + D_j - 2w_{ij}$

if $J_{ij} > 0$ ⇒ improve

Girvan Newman Alg *

1 edge betweeness number of shortest Path passing over the edge
high ⇒ connect different clusters

1 - compute betweeness → detail in slide
2 - remove highest betweeness value
3 - recompute betweeness
until graph have K-components
?
Modularity Q

Modularity Q

$Q \sim$

Collective classification — node content + graph structure

↓

Classify node

ICA (iterative classification alg)

link feature — | - incident node
- Structural properties of graph (page rank, ...)

generate

content feature

training (iteratively)

Book Reg: Mining massive datasets

Master data science HUS, VNU cinnamon

ngoài giờ

T5 - 2018