

16-06-2018: Social network analysis

- Hong Le Hong Phuong -
HUS, VNU
chief scientist of FPT

Network ~~model~~ → cluster → communities

SNAP - standard
social network analysis platform

Homophily: same properties → connect

Triadic closure:

Clustering coefficient

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

Measure of centrality:

prestige
closeness centrality

proximity prestige
betweenness centrality for node
edge

COMMUNITY DETECTION

[spectral clustering]*
structural properties of network

"spectral representation"
embed nodes into a multidimensional
space \mathbb{R}^d

node i → \vec{v}_i d-dimensions
 \mathbb{R}^d

eigen vector
value

"spectrum" of matrix

matrix representation (adjacency matrix) A
degree matrix D
laplacian matrix $L = D - A$

x^T $|$

d=1

cinnamon

nodes $\rightarrow y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n$
node n

Edge weight $W = (w_{ij})_{i,j \in V}$

minimise $J(y) = \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - y_j)^2$
over $S(y)$

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

diagonal matrix $\Lambda = \begin{pmatrix} \Delta_{11} & & \\ & \ddots & \\ & & \Delta_{nn} \end{pmatrix}$

$L = \Lambda - W$

rewrite $\Rightarrow J(y) = 2y^T L y$

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

\rightarrow solve optimisation problem

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

Kernighan-Lin alg

heuristic
NP-hard
local optima

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 $J_{ij} > 0 \Rightarrow$ improve

Girvan-Newman Alg

- 1 - edge betweenness number of shortest paths passing over the edge
- high \Rightarrow connect different clusters

- 1 - compute betweenness
 - 2 - remove highest betweenness value
 - 3 - recompute betweenness
- until graph has K components

Modularity Q

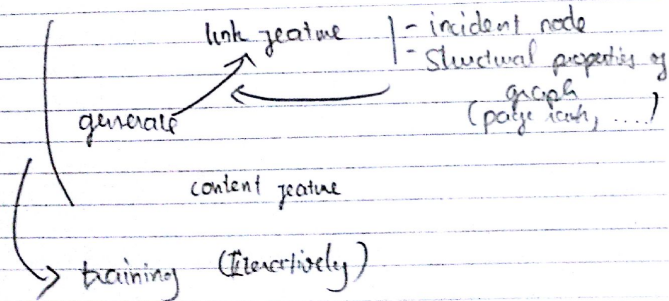
Modularity &

$Q \sim$



Collective classification
+ node content + graph structure
↓
classify node

ICA (iterative classification alg)



Book Regs. Using massive dataset

Master Data Science
ngoai yid
T5 - 2018

HUS, VNU



cinnamon