PAKDD2020
SINGAPORE

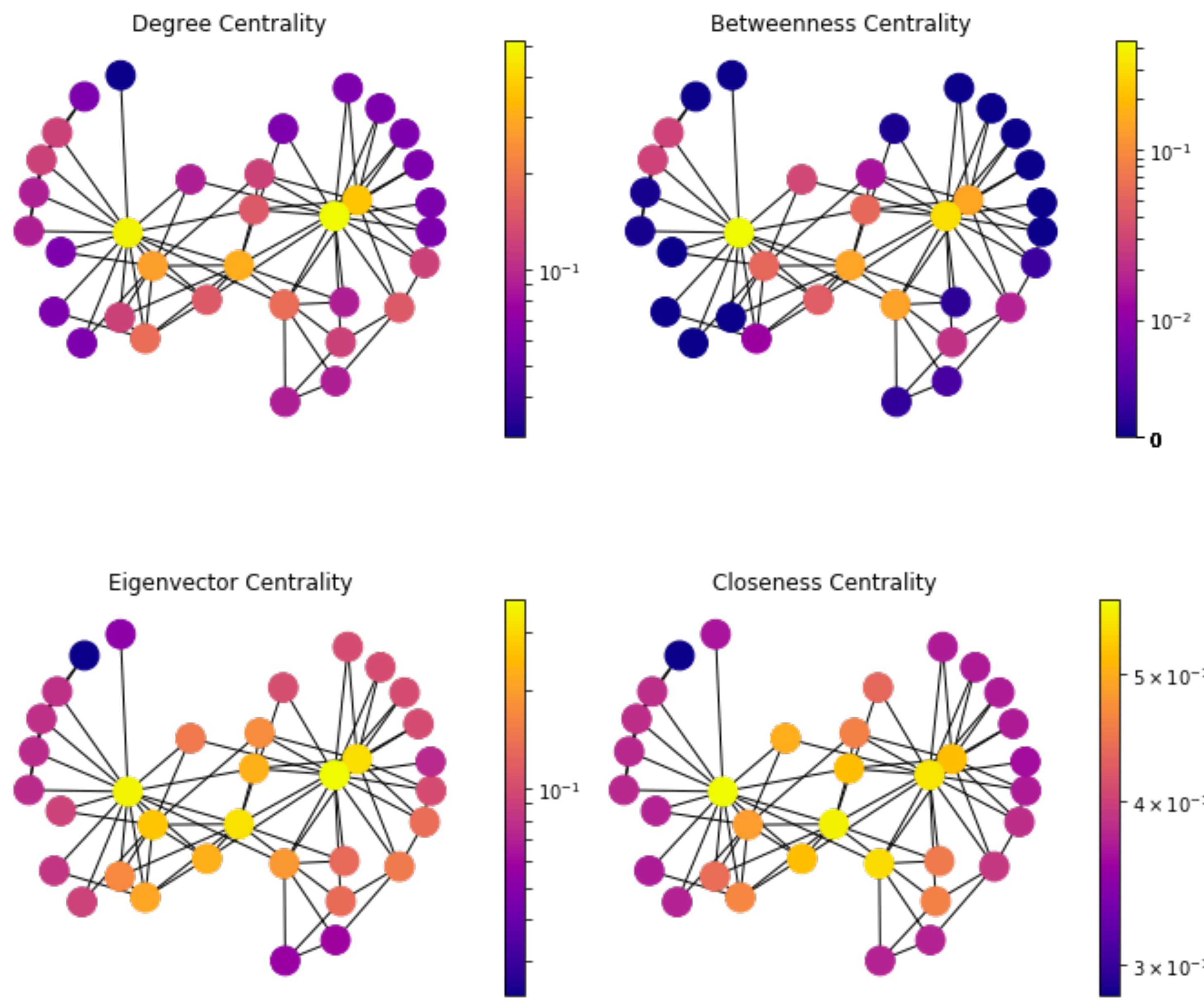
北京大學
PEKING UNIVERSITY

NODE CONDUCTANCE:
A Scalable Node Centrality Measure on Big Networks

Tianshu Lyu, Fei Sun, Yan Zhang
Peking University, Alibaba Group

Node Centrality

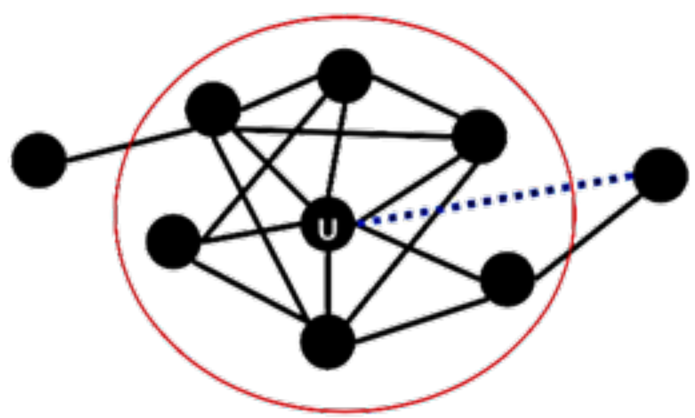
- **Local Centrality**
  - ego-network
  - less informative
- **Global Centrality**
  - ideal routes
  - unrealistic
  - infeasible to compute



Refer to https://aksakalli.github.io/2017/07/17/network-centrality-measures-and-their-visualization.html

Node Conductance

- **Conductance**
  - measures how hard it is to leave **a set of** nodes



$$\phi(S) = \frac{\mathrm{cut}(S)}{\min(\mathrm{vol}(S), \mathrm{vol}(\bar{S}))}$$

- **Node Conductance**
  - measures how hard it is to leave **a certain** node
  - **the sum of the probability that i is revisited at s-th step, where s is the integer between 1 and ∞.**

$$\mathtt{NC}_{\infty}(i) \equiv \sum_{s=1}^{\infty} P(i|i, s).$$

Node Conductance

$$\mathtt{NC}_{\infty}(i) \equiv \sum_{s=1}^{\infty} P(i|i,s).$$

Definition	Notation
adjacency matrix	**A**
degree vector	**d** = **A1**
diagonal matrix of degree	**D** = diag(**d**)

- For a walk starting at node i, the probability that we find it at j after exactly s steps is given by

$$P(j|i,s) = [(\mathbf{D}^{-1}\mathbf{A})^s]_{ij}.$$

- NC_r denotes the sum of the probability that the node is revisited at the step s, s is between 1 and r

$$\mathtt{NC}_r(i) = \Sigma_{s=1}^{r} P(i|i,s) = \mathbf{P}_{ii}^{(r)}, \quad \mathbf{P}^{(r)} = \Sigma_{s=1}^{r}(\mathbf{D}^{-1}\mathbf{A})^s,$$

Node Conductance

- Supposed that r approaches infinity,

$$\mathbf{P}^{(\infty)} = \Sigma_{s=1}^{\infty}(\mathbf{D}^{-1}\mathbf{A})^s = \Sigma_{s=0}^{\infty}(\mathbf{D}^{-1}\mathbf{A})^s - \mathbf{I}$$
$$= (\mathbf{I}-\mathbf{D}^{-1}\mathbf{A})^{-1} - \mathbf{I} = (\mathbf{D}-\mathbf{A})^{-1}\mathbf{D} - \mathbf{I}.$$

Laplacian Matrix

- Pseudo-inverse of Laplacian Matrix:

$$g(\lambda_k) = \begin{cases} \frac{1}{\lambda_k}, & \text{if} \quad \lambda_k \neq 0 \\ 0, & \text{if} \quad \lambda_k = 0 \end{cases}, \quad \mathbf{L}_{ii}^{\dagger} = \Sigma_{k=1}^{N-1} g(\lambda_k) u_{ik}^2,$$

- Node Conductance only concerns about the diagonal:

$$\texttt{NC}_{\infty}(i) \propto \mathbf{L}_{ii}^{\dagger} \cdot d_i,$$

Related Work

$$\mathrm{SC}(i) = \sum_{s=1}^{\infty} \frac{(\mathbf{A^s})_{\mathbf{ii}}}{s!}$$

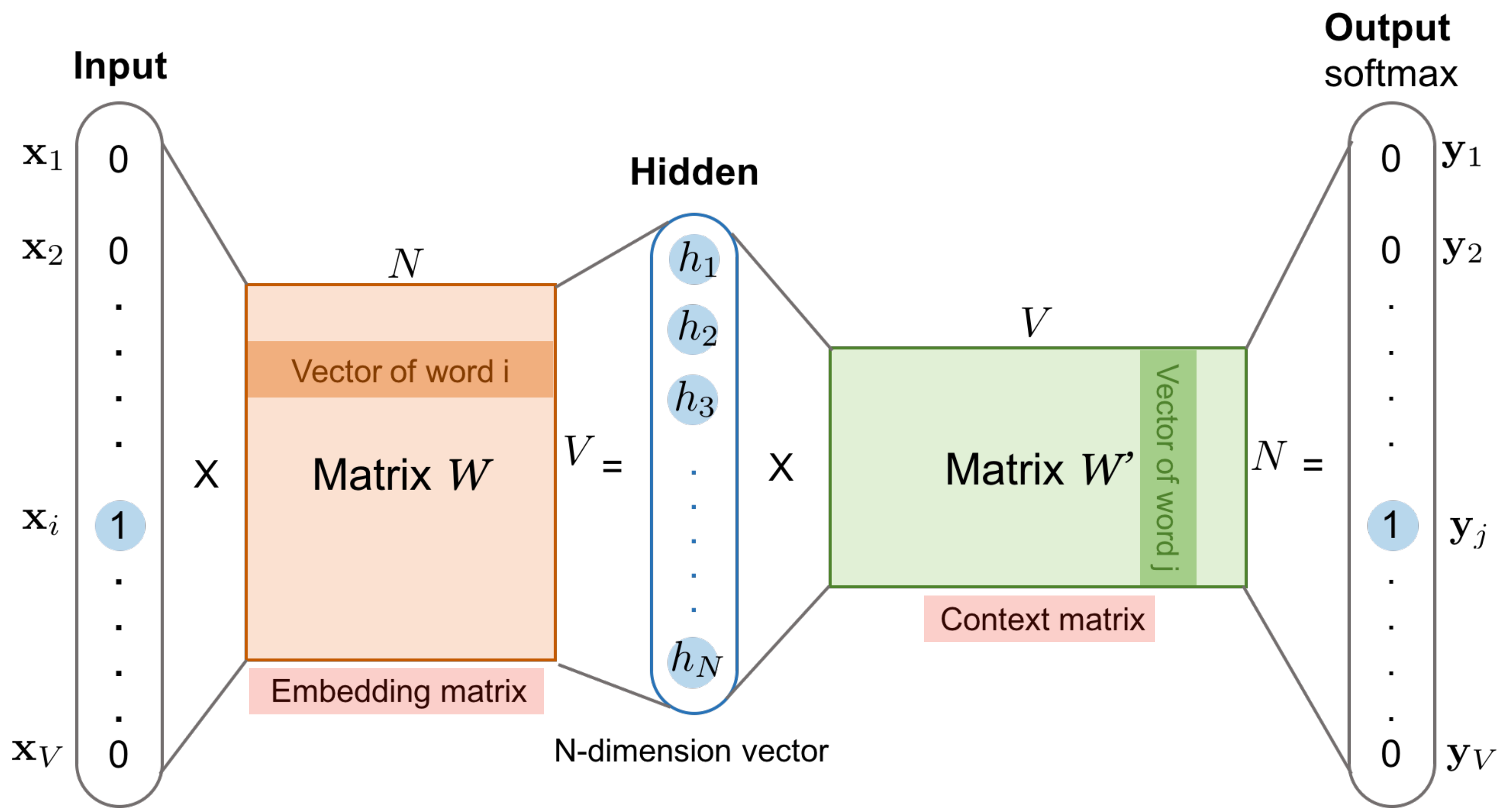
- Subgraph Centrality
  - the “sum” of closed walks of different lengths in the network starting and ending at vertex i.
  - adding a scaling factor to the denominator in order to make the SC value converge
  - less interpretive

Related Work

$$\mathrm{PR} = \mathbf{D}(\mathbf{D} - \alpha\mathbf{A})^{-\mathbf{1}}\mathbf{1}$$

- PageRank
  - the stationary distribution of the random walk
  - the probability that a random walk, with infinite steps, starts from **any node** and hits the node under consideration.

DeepWalk Vectors



- **Syntagmatic:** If two nodes have a strong connection in the network, the value of $\mathbf{w}_i \cdot \mathbf{c}_j$ is large.

- **Paradigmatic:** If two nodes have similar neighbors, the value of $\mathbf{w}_i \cdot \mathbf{w}_j$ is high.

DeepWalk Vectors & Node Conductance

- DeepWalk loss function

$$\mathcal{L} = \sum_{i \in \mathcal{V}_W} \sum_{j \in \mathcal{V}_C} \#(i,j)_r \left(\log \sigma(\mathbf{w}_i \cdot \mathbf{c}_j) \right) + \sum_{i \in \mathcal{V}_W} \#(i)_r \left(k \cdot \sum_{\text{neg} \in \mathcal{V}_C} P(\text{neg}) \log \sigma(-\mathbf{w}_i \cdot \mathbf{c}_{\text{neg}}) \right).$$

- Comparing the derivative to zero

$$\mathtt{NC}_r(i)$$

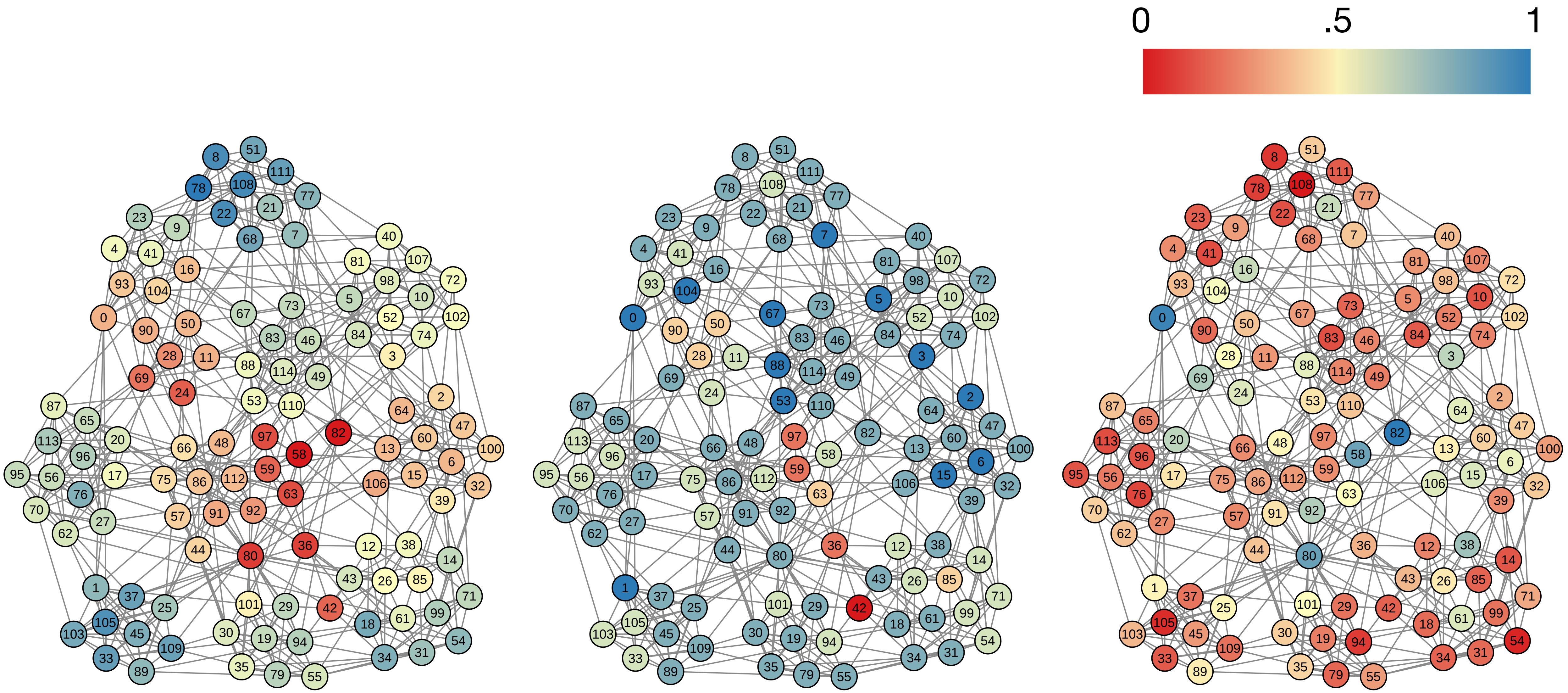
$$\mathbf{w}_i \cdot \mathbf{c}_j = \log\left(\frac{\#(i,j)_r}{\#(i)_r \cdot P(j)}\right) - \log k,$$

- Estimating probability by the actual number of observations

$$\mathtt{NC}_r(i) = \exp(\mathbf{w}_i \cdot \mathbf{c}_i) \cdot k \cdot P(i) \propto \exp(\mathbf{w}_i \cdot \mathbf{c}_i) \cdot \deg(i).$$

Compute approximate Node Conductance by DeepWalk vectors

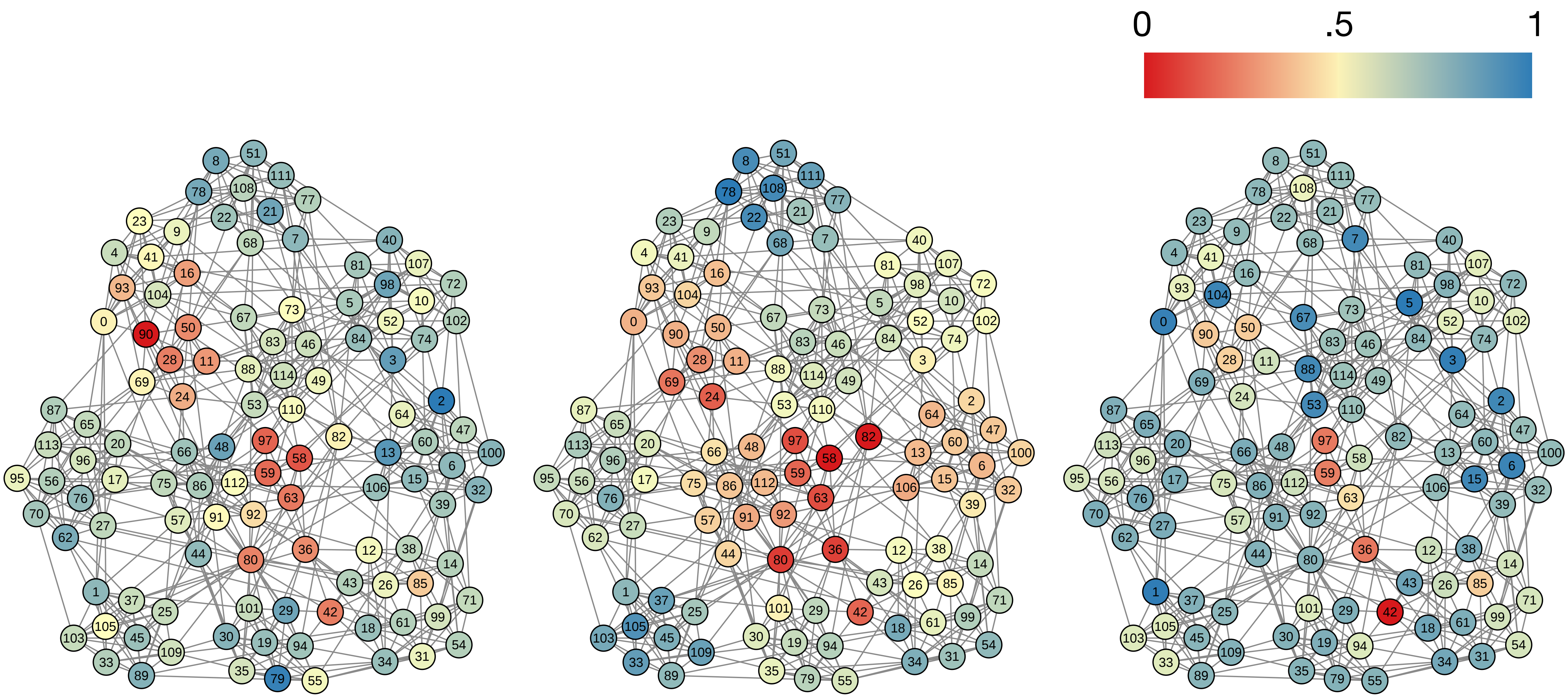
Visualization



Node Conductance Degree Betweenness

Node Conductance gives low value to nodes with low degree and high betweenness centrality.

Visualization

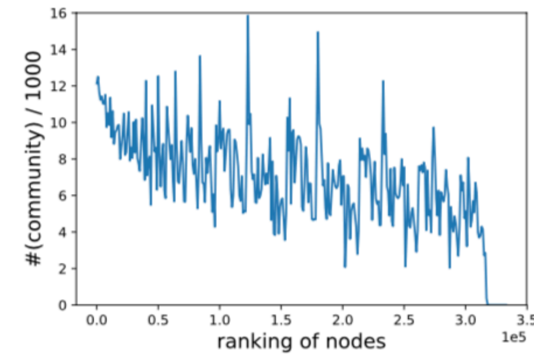


Node Conductance (DW) | Node Conductance | PageRank

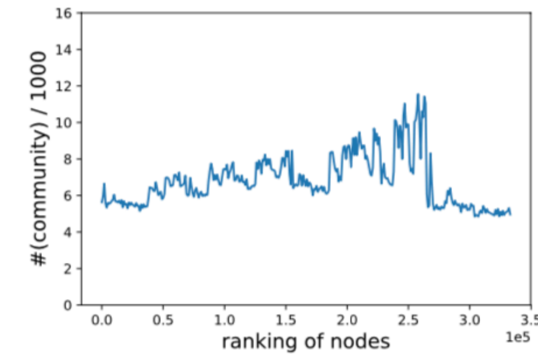
Node Conductance looks quite different with PageRank. Node Conductance and its approximation are similar.

Finding Nodes Spanning Several Communities

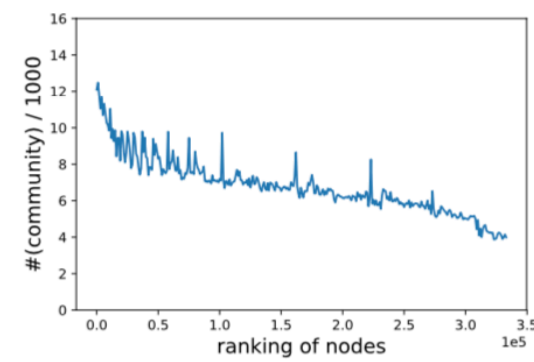
- Plot the communities numbers of nodes (y-axis) in the order of each centrality measure (x-axis).
- Node Conductance provides the smoothest curve comparing with the other five metrics.



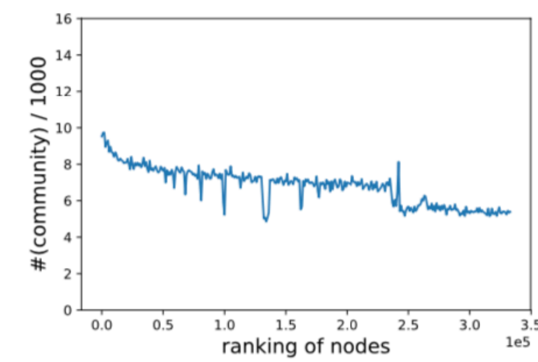
(a) Degree Centrality



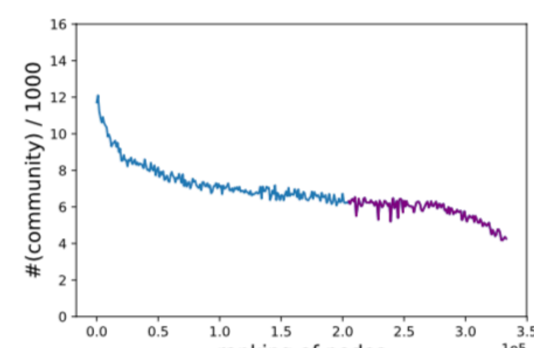
(b) Clustering Coefficient



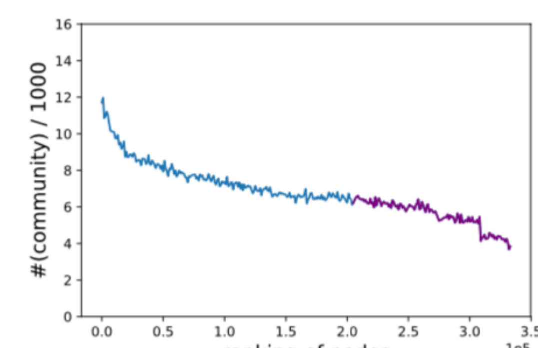
(c) Eigenvector Centrality



(d) Approximate Betweenness



(e) PageRank



(f) Node Conductance

Conclusion

- Intuition
  - the probability of revisiting the target node in a random walk
- Approximation
  - by the dot product of the input and output vectors
- Mining influential nodes

PAKDD2020
SINGAPORE

北京大学
PEKING UNIVERSITY

THANKS!

Tianshu Lyu, Fei Sun, Yan Zhang
Peking University, Alibaba Group