Graph Analytics

Jay Pujara

DSCI558 Spring 2021

3/31/21

Illustrations from A Comprehensive Guide to Graph Algorithms in Neo4J

The most useful graph algorithms

- Pathfinding and Search
- Matching and coloring
- Centrality
- Clustering and Community Detection
- Partitioning

The most useful graph algorithms

- Pathfinding and Search
  - Djikstra's algorithm
- Matching and coloring
  - B-matching, k-coloring, graph equivalence
- **Centrality**
- **Clustering and Community Detection**
- Partitioning
  - Minimum weighted edge cut (METIS)

Basic Concepts

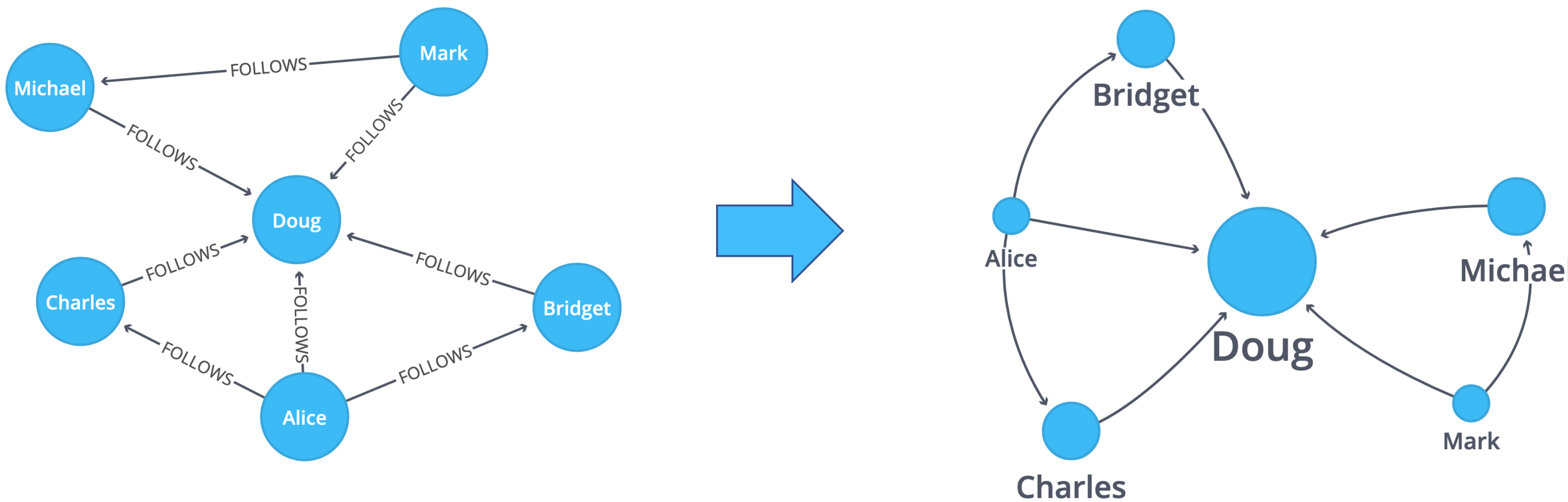
- **Eccentricity**: Maximum distance of a node to other nodes in the graph
- **Central point**: node with lowest eccentricity
- **Radius**: distance: central point’s eccentricity (lowest max-distance)
- **Diameter**: largest eccentricity (highest max-distance)

Centrality

- Degree centrality
- Eigenvector Centrality / PageRank / Personalized PageRank
- Betweenness centrality
- Closeness (Harmonic) centrality
- HITS (hyperlink-induced topic search / hubs & authorities)
- GENI

Degree Centrality

- Nodes with maximum in-degree or out-degree
- Normalized my maximum degree.



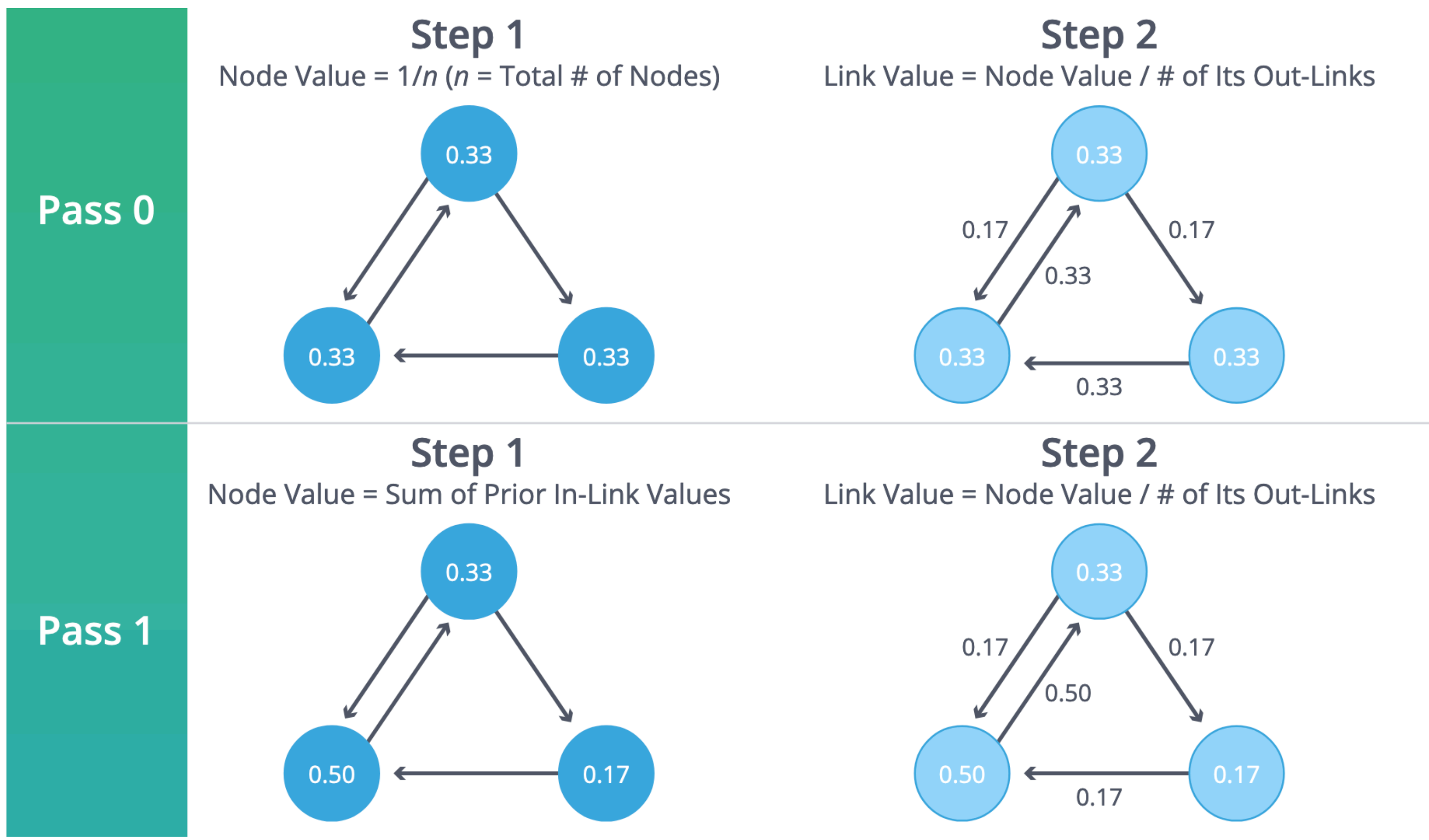
PageRank

- Probability a random walk (with restart) will reach a node

- Iterative algorithm, each iteration distributes node's PageRank to neighbors

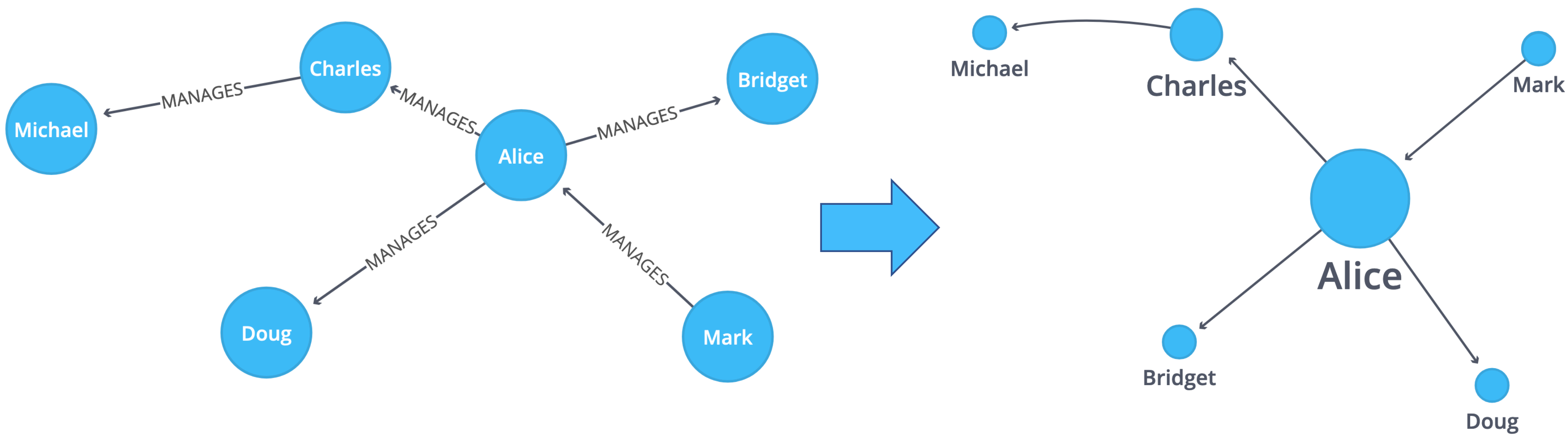
- PR(A) = 1-d + d sum(PR(V_1)/OD_1, PR(V_2)/OD_2, PR(V_n)/OD_n)

PageRank illustrated



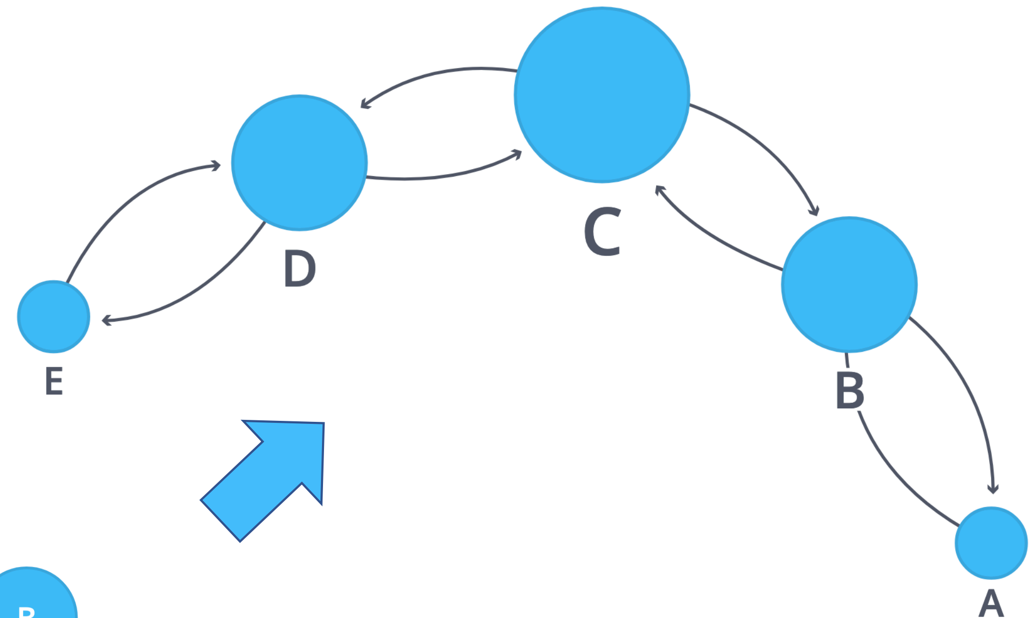
Betweenness Centrality

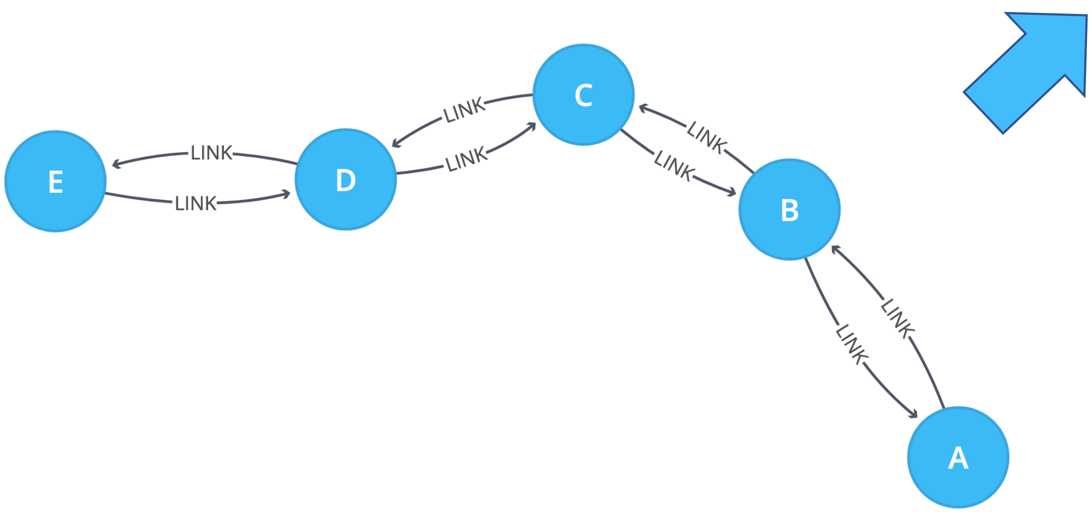
- Number of all-pairs shortest paths a node participates in



Closeness (Harmonic) Centrality

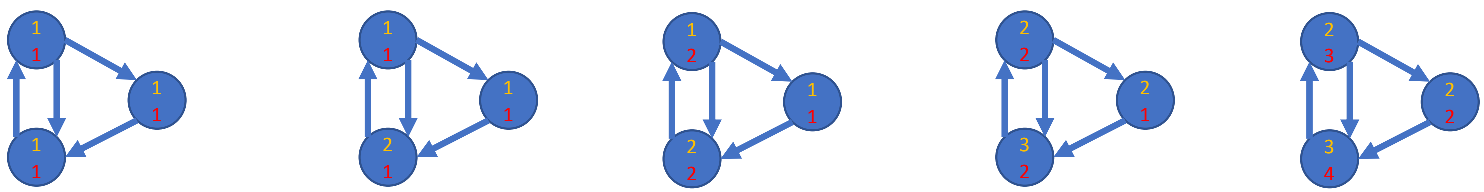
- Inverse average of distance (or average inverse distance) to all other nodes





HITS (hubs & authorities)

- Initialize hub / authority to 1 for all nodes
- Update authority score to sum of in-edges associated hub scores
- Update hub score to sum of in-edges associated authority scores
- Normalize*

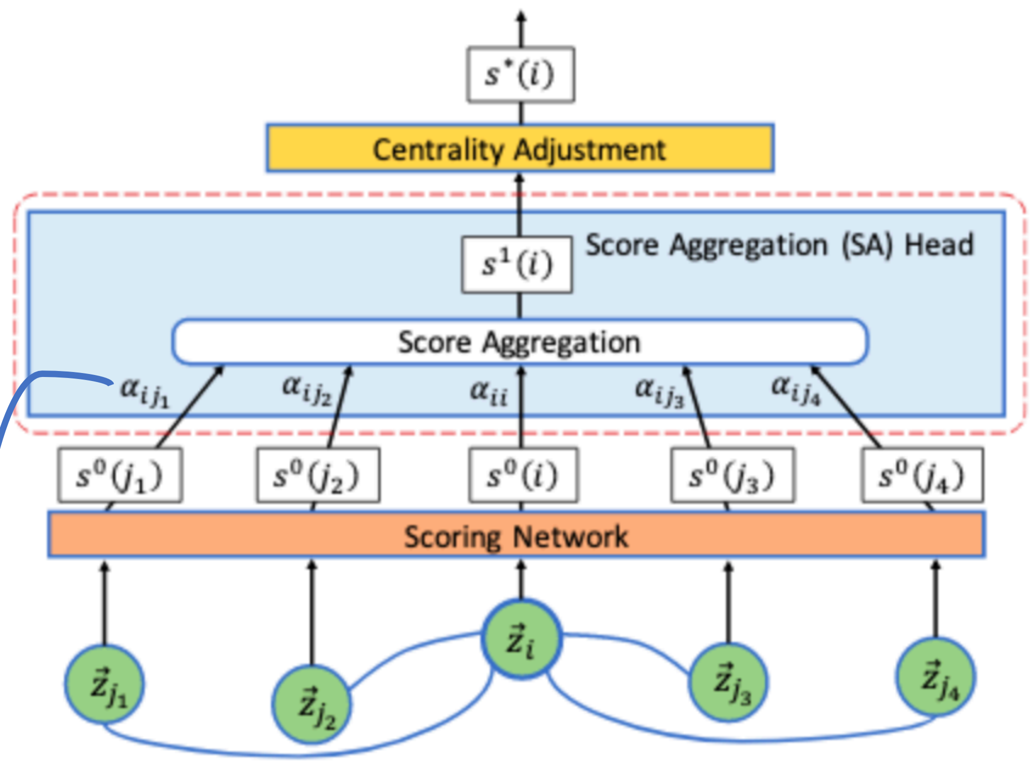


GENI

Estimating Node Importance in Knowledge Graphs Using Graph Neural Networks

Symbol	Definition
V_s	set of nodes with known importance scores
$\vec{z}_i$	real-valued feature vector of node i
$\mathcal{N}(i)$	neighbors of node i
L	total number of score aggregation (SA) layers
ℓ	index for an SA layer
H^ℓ	number of SA heads in ℓ-th layer
p_{ij}^m	predicate of m-th edge between nodes i and j
$\phi(e)$	learnable embedding of predicate e
σ_a, σ_s	non-linearities for attention computation and score estimation
$s_h^\ell(i)$	estimated score of node i by h-th SA head in ℓ-th layer
$s^*(i)$	centrality-adjusted score estimation of node i
$\|\|$	concatenation operator
$d(i)$	in-degree of node i
$c(i)$	centrality score of node i
$c_h^*(i)$	centrality score of node i scaled and shifted by h-th SA head
γ_h, β_h	learnable scale and shift parameters used by h-th SA head
$\vec{a}_{h,\ell}$	learnable parameter vector to compute $\alpha_{ij}^{h,\ell}$ by h-th SA head in ℓ-th layer
$\alpha_{ij}^{h,\ell}$	node i's attention on node j computed with h-th SA head in ℓ-th layer
$g(i)$	known importance score of node i

Use local node importance features and node / neighbor embeddings



$$\alpha_{ij}^\ell = \frac{\exp\left(\sigma_a\left(\sum_m \vec{a}_\ell^\top [s^\ell(i) \| \phi(p_{ij}^m) \| s^\ell(j)]\right)\right)}{\sum_{k\in\mathcal{N}(i)\cup\{i\}} \exp\left(\sigma_a\left(\sum_m \vec{a}_\ell^\top [s^\ell(i) \| \phi(p_{ik}^m) \| s^\ell(k)]\right)\right)}$$

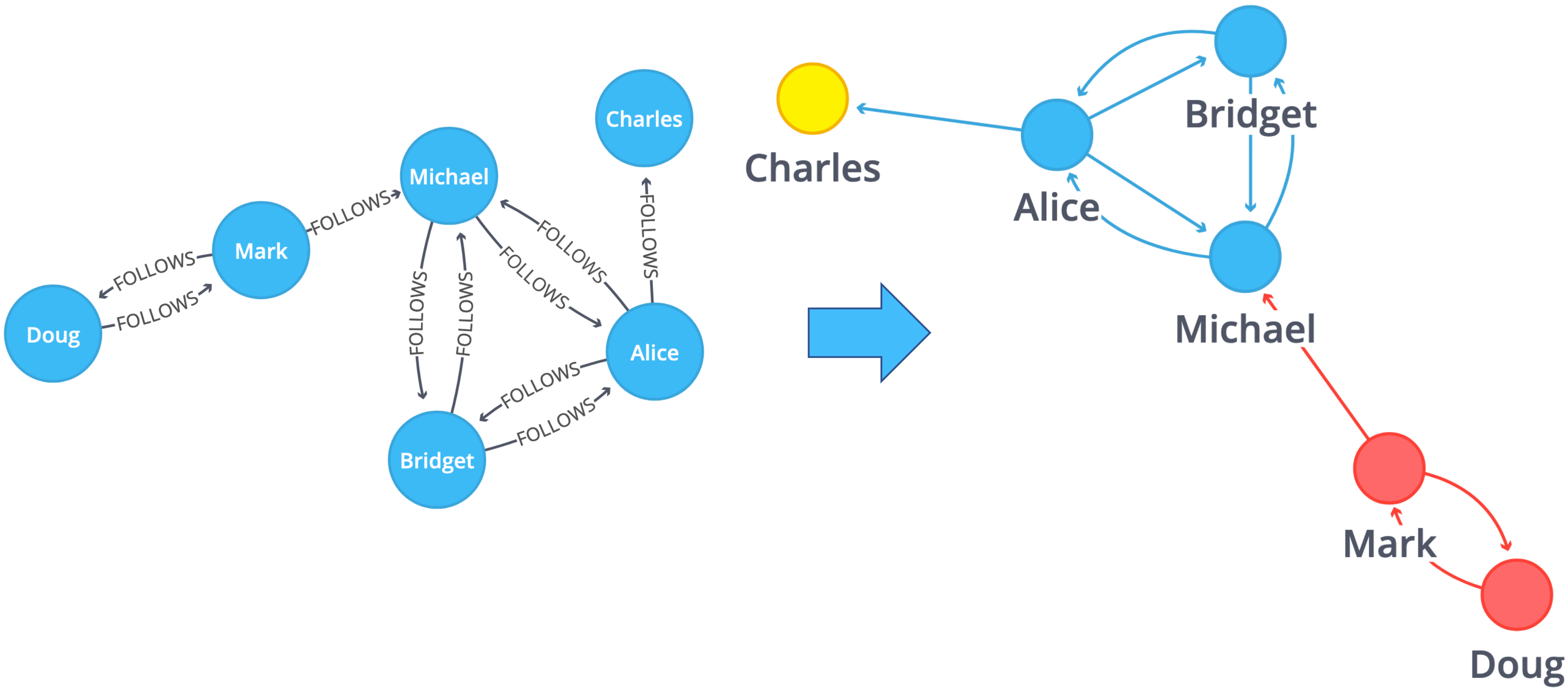
attention depends on predicates

Community Detection

- Connected components
- Modularity
- Clustering coefficient

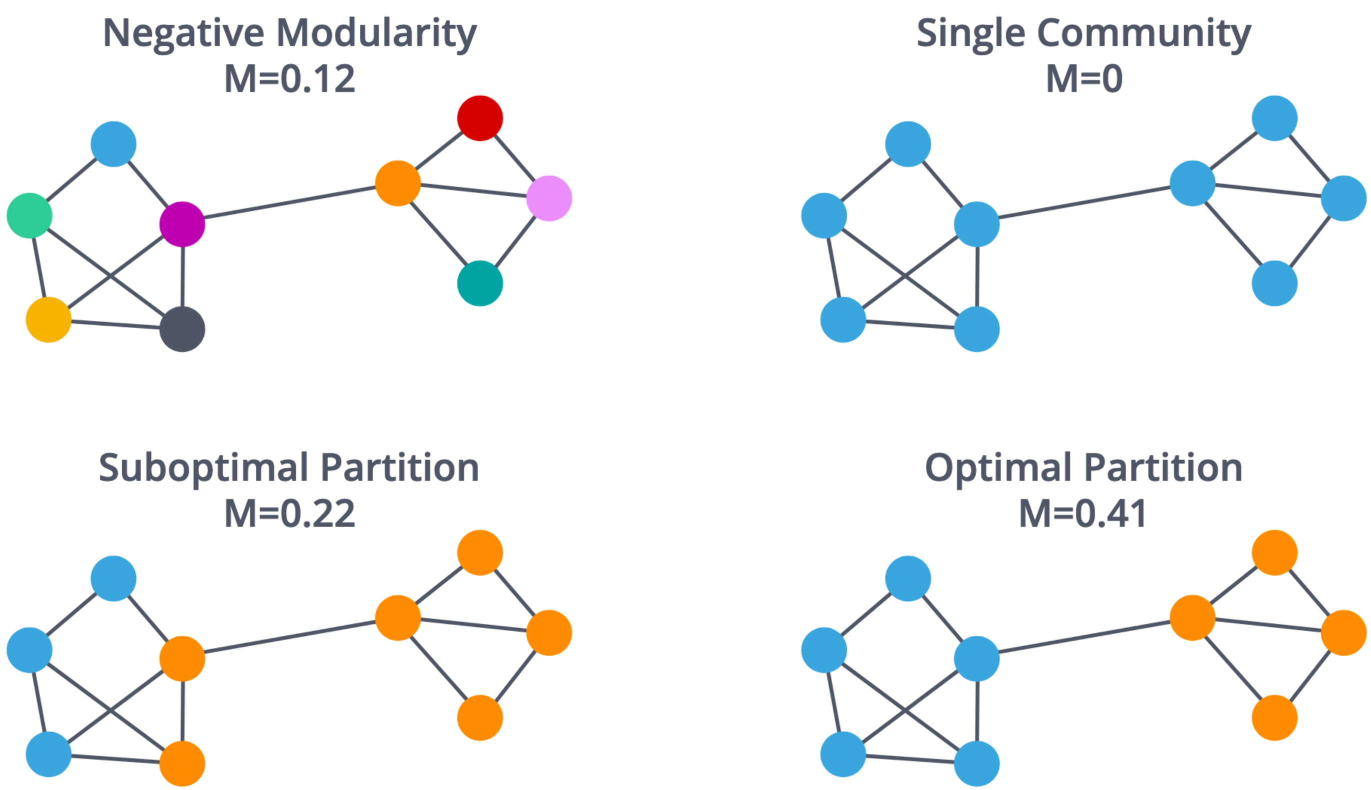
Connected components

- Set of nodes where each node is reachable from all other nodes



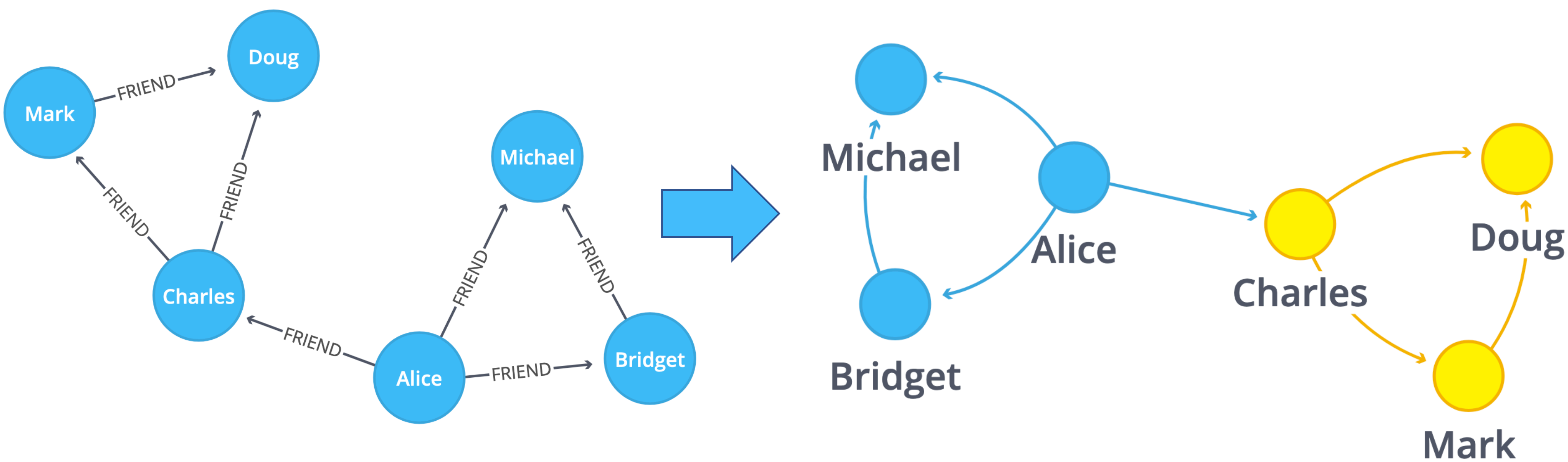
Modularity

- Number of within-cluster edges minus expected edges in random graph
- Basic random graph: cut each edge in half, rewire to other half-edge



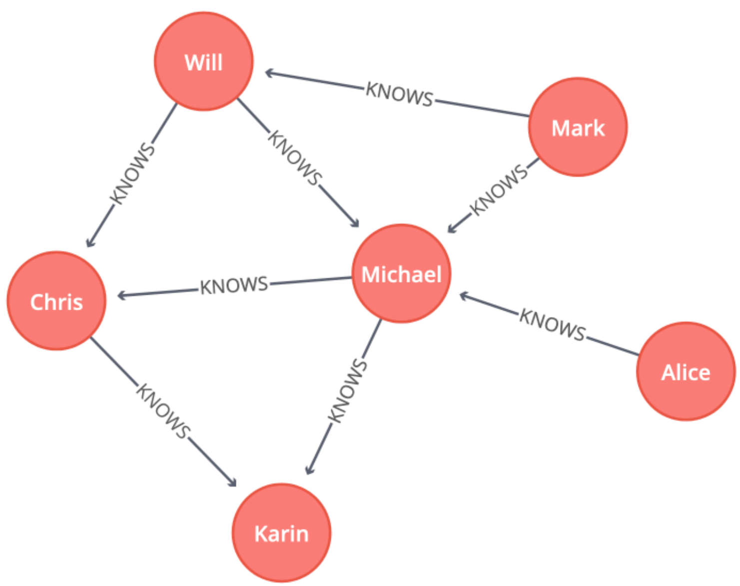
Modularity

- Number of within-cluster edges minus expected edges in random graph
- Basic random graph: cut each edge in half, rewire to other half-edge



Clustering Coefficient

- Fraction of neighbors who are connected to each other
- Computed via number of closed triangles over all possible triangles



Additional Resources

- David Kempe's "Structure and Dynamics of Information in Networks"
- Course on Structured Analysis and Visualization of Networks
- Neo4J manual on Graph Algorithms
- NetworkX and SNAP
- Slides on Social Network Analysis