Scalable Graphs

Graphs and graph analytics with GraphLab Create

Why network analysis?

To understand relationships between entities

- Key nodes
- Outliers and anomalies
- Segmentation
- Connections between nodes of interest
- Comparisons between networks
- Visualization

Scalable Graph = SGraph

SGraph

- On-disk graph store
- Backed by SFrames
- Simple attribute creation
- Fast, straightforward queries
- Lazy evaluation and immutability for chaining operations
- Flexible and powerful computation with triple apply

Basic SGraph usage

- Demo time!
- Recap
  - Constructed an SGraph from scratch
  - Summarized and visualized
  - Attribute manipulation
  - Retrieved vertex neighborhoods
  - Triple apply for more sophisticated computation

Graph Analytics

Tasks vs. tools

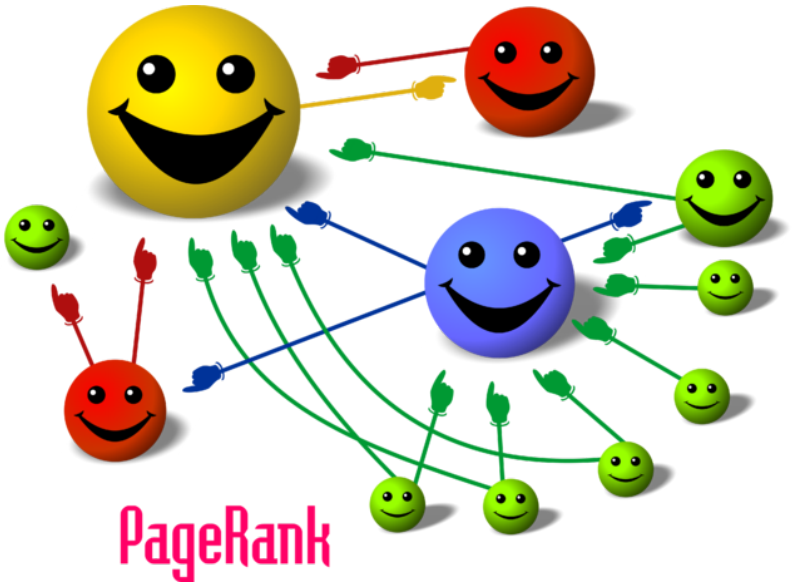
	Degree	Page rank	Triangle count	K-core	Com-ponents	Coloring	Shortest path
Key nodes	✓	✓	✓	✓			✓
Outliers	✓	✓	✓	✓	✓		
Clusters				✓	✓	✓	
Hidden Connections					✓		✓
Network comparisons	✓	✓	✓	✓	✓	✓	

Tasks vs. tools

	Degree	Page rank	Triangle count	K-core	Com-ponents	Coloring	Shortest path
Key nodes	✓	✓	✓	✓			✓
Outliers	✓	✓	✓	✓	✓		
Clusters				✓	✓	✓	
Hidden Connections					✓		✓
Network comparisons	✓	✓	✓	✓	✓	✓	

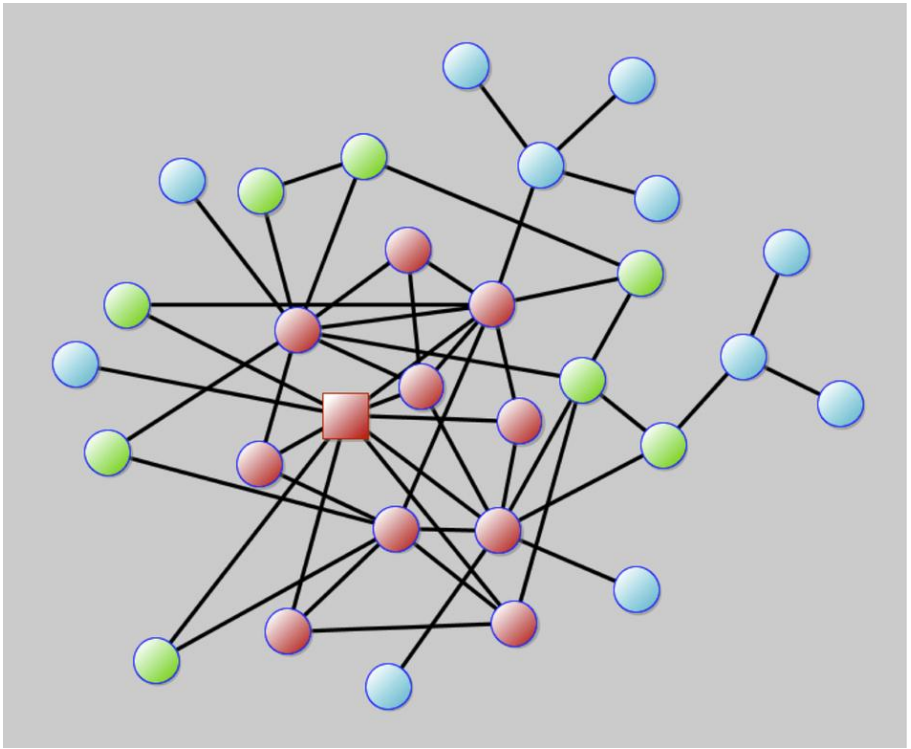
Finding key nodes - pagerank

- How influential am I?
- Weighted sum of the influence of nodes directed toward me.
- Compute iteratively.



Finding key nodes - k-core

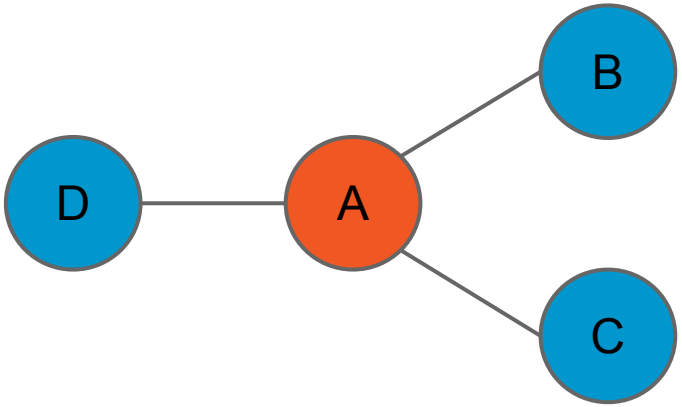
- How close to the graph "core" am I?
- Recursively prune nodes with degree *k*.

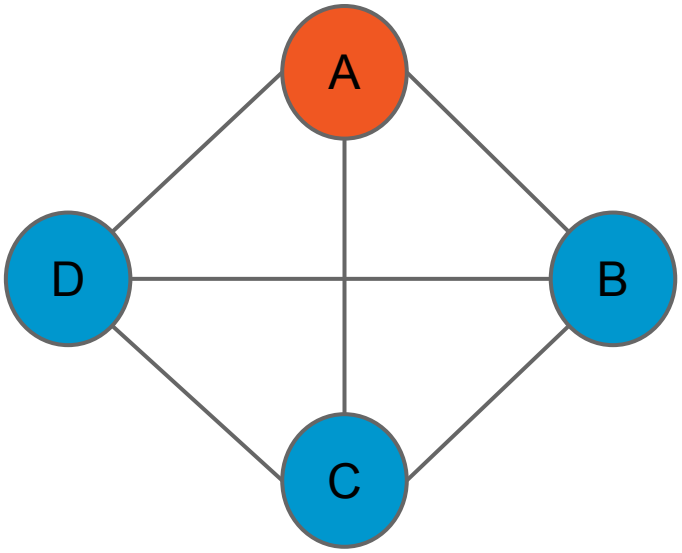


Castellano, C. and Pastor-Satorras, R. (2012) "Competing activation mechanisms in epidemics on networks." *Scientific Reports* 2: 371. http://www.nature.com/srep/2012/120420/srep00371/fig_tab/srep00371_F1.html

Finding key nodes - triangle count

- How dense is my network?
- Number of my pairs of friends who are friends with each other.





Graph analytics usage

- Demo time!

Exercises