

Interpreting the results of community detection algorithms

Nathan Smith,
Senior data scientist

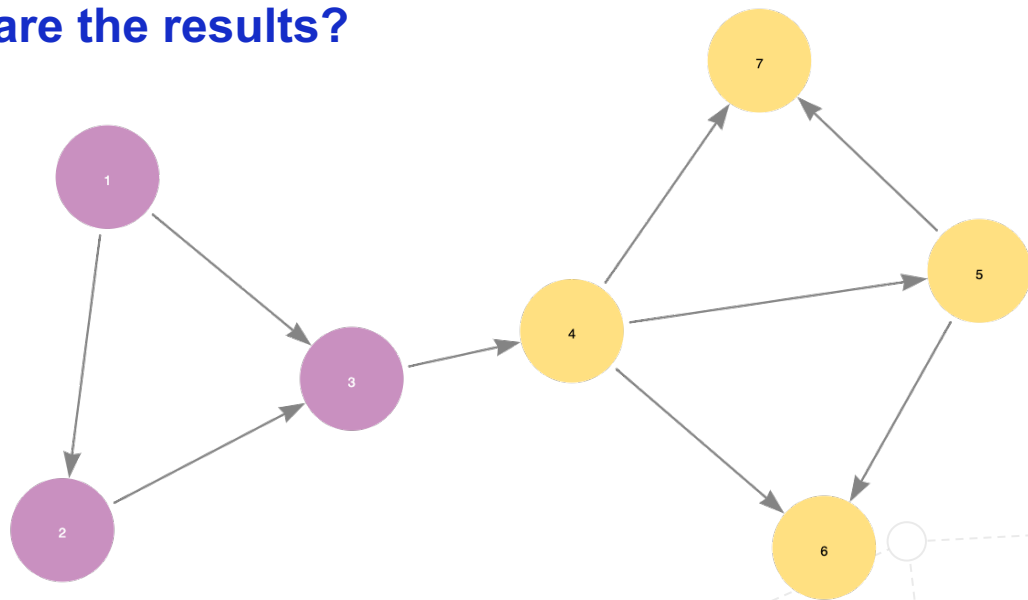


Sample code repo

Congratulations!

You have successfully run a community detection algorithm!

But how useful are the results?



What makes a good community?

- Weak relationships to other communities
- Strong relationships within the community
- Identifiable characteristics of community members

We can measure community quality!

- Weak relationships to other communities

Conductance and
Modularity

- Strong relationships within the community

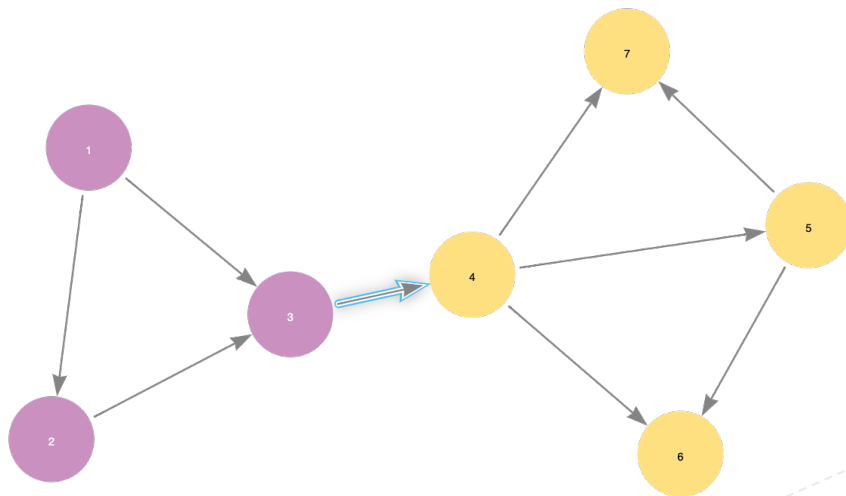
Clustering coefficient

- Identifiable characteristics of community members

Centrality

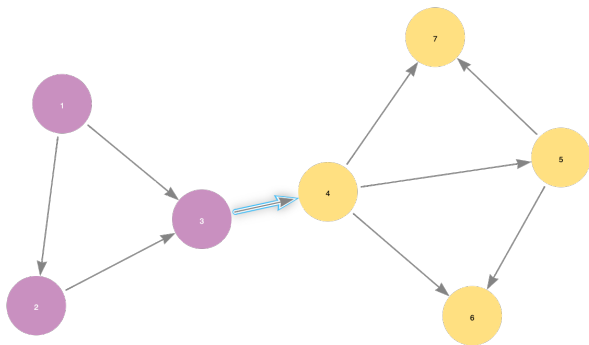
Conductance

- What percentage of relationships that start in a community end in the same community?
- Lower conductance scores mean more distinct communities.

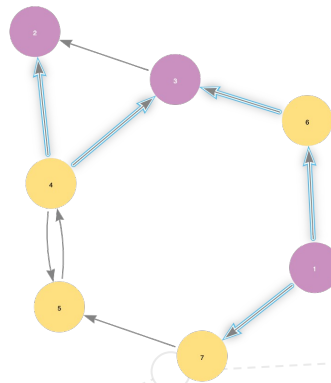


Modularity

- What is the difference between the ratio of relationships with both endpoints in a community compared to what the ratio would be if the relationships were distributed randomly?
- Higher modularity scores mean more distinct clusters.



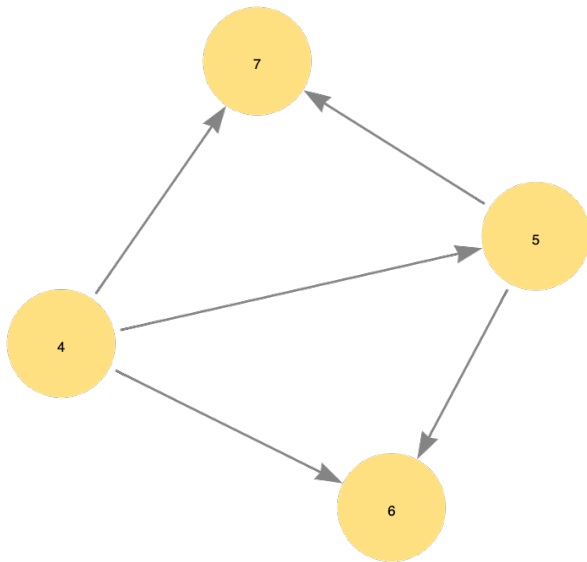
Original relationships



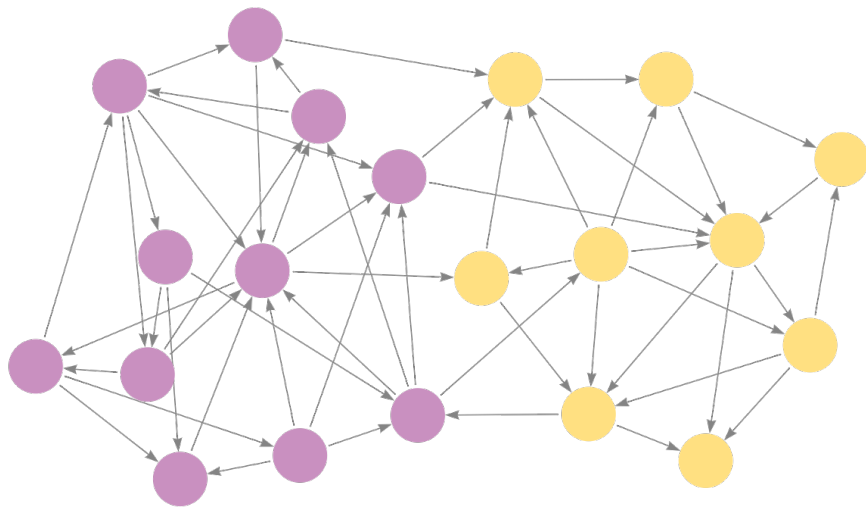
Randomly reassigned relationships

Clustering coefficient

- What percentage of the neighbors of a node are related to each other?
- Higher scores mean more connected, cohesive clusters

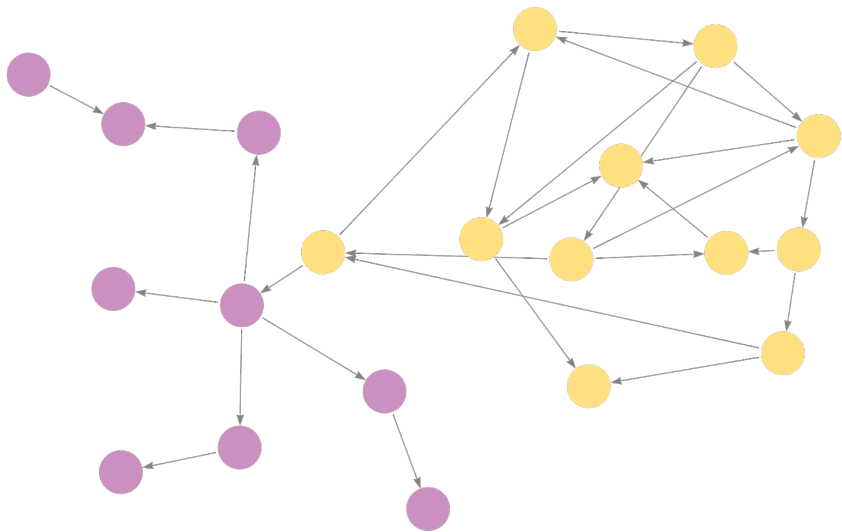


Example graph



- **Conductance:**
 - Purple: 0.10
 - Yellow: 0.13
- **Modularity: 0.38**
- **Clustering Coefficient:**
 - Purple: 0.54
 - Yellow: 0.65

Example graph



- **Conductance:**
 - Purple: 0.05
 - Yellow: 0.02
- **Modularity: 0.38**
- **Clustering Coefficient:**
 - Purple: 0.00
 - Yellow: 0.12

Demo



https://github.com/smithna/blogs/tree/main/community_quality

What to tell your boss

- Are the communities distinct and cohesive enough to be useful?
- Are the community quality statistics changing over time?
- What are high centrality examples within each community?
- Thank you for sending me to GraphConnect!

Thank you!

Contact us at
nathan.smith@neo4j.com
@nsmith_piano on Twitter

