Network Centrality

6/14/2022

- worksheet:
  - If you haven't completed both problems from yesterday, you don't have to turn it in yes
  - Turn in by Friday

- Tips:
  - work with someone else! These problems are not easy so find someone to bounce ideas off of.
  - Do it by hand first
  - Come to office hours
  - write pseudocode before you code

pipeline for solving these problems: make sure you can do it by hand

↓

write pseudocode

↓

code it up

I have uploaded two resources on the website:

- networks textbook pages 1-25
- numpy reference

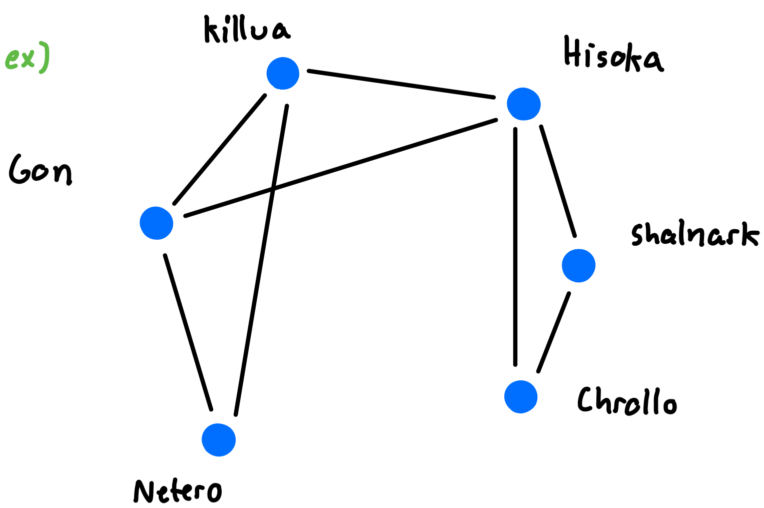
Office hours this week and going forwards: Th 1-3 in Rhodes 657

Network Centrality

one way to analyze a network.

Network centrality measures which nodes are important in a graph
What does this mean? Depends on the context of the application.

ex)



Show called "Hunter x Hunter"

we might look for the node (or character in this case) that has the largest degree: Hisoka

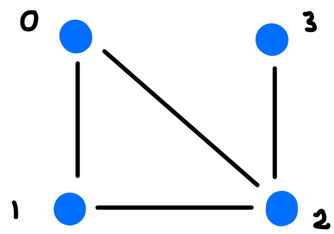
But there are many different ways to define node importance so let's explore a few

Network Centrality Methods:

① Degree centrality
idea: a node is important if it has many neighbors (if it has high degree)

Steps: • Compute the degree matrix
• Find the nodes with the largest degrees

ex)



Recall the degree matrix stores the degree of each node in a network

$$D = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Node with the largest degree centrality is node 2.

Here this was a very simple network where you didn't really need to compute the degree matrix when you work with adjacency lists you would need to compute the degree matrix D.

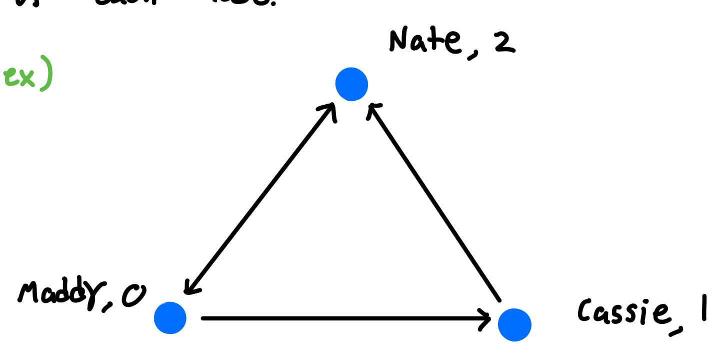
② Eigenvector Centrality
idea: a node with 300 non-influential friends is much less important than someone with 300 influential friends (like Barack Obama)

degree centrality would label them as having equal influence but eigenvector centrality would label Obama as being more important.

Steps: • compute the adjacency matrix of the network
• compute the eigenvalues of the adjacency matrix
• compute the eigenvector associated with the largest $|\lambda|$
• Normalize the eigenvector

Components of the normalized eigenvector correspond to the importance of each node.

ex)



$$A = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

$\lambda = 1.325,\ -0.66 - 0.56i,\ -0.66 + 0.56i$

The largest $|\lambda| = 1.325$

For the largest $|\lambda|$, the eigenvector is:

$$v = \begin{bmatrix} 1.325 \\ 0.755 \\ 1 \end{bmatrix} \xrightarrow{\text{normalize}} v = \begin{bmatrix} 0.727 \\ 0.414 \\ 0.548 \end{bmatrix} \begin{array}{l} \leftarrow \text{Maddy's influence} \\ \leftarrow \text{Cassie's influence} \\ \leftarrow \text{Nate's influence} \end{array}$$

For larger networks, computing the eigenvalues and eigenvectors of the adjacency matrix becomes computationally infeasible.

So we use the power iteration method to find the eigenvector associated with the largest absolute value eigenvalue.

Power iteration method: computational method for computing the eigenvector associated with the largest $|\lambda|$

Steps:
- Let $x^{(0)}$ be an arbitrary vecter of length n for a graph of n nodes.
- Repeated compute $x^{(t)} = \dfrac{Ax^{(t-1)}}{\|Ax^{(t-1)}\|}$ until $t = T$

Side note: ex) 3 nodes in the graph

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad x^{(0)} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$x^{(1)} = \frac{A \cdot x^{(0)}}{\|A \cdot x^{(0)}\|}$$

$$x^{(2)} = \frac{A \cdot x^{(1)}}{\|A \cdot x^{(1)}\|}$$

$$x^{(3)} = \frac{A \cdot x^{(2)}}{\|A \cdot x^{(2)}\|}$$

$$\vdots$$

- associate each entry in $x^{(T)}$ to it's corresponding node

Note: T is arbitrary. You can choose T by looking at $x^{(t)}$ and stopping doesnt change that much any more.

You can either preliminarily set T or you can see how much $x^{(t)}$ changes and stop the loop when $x^{(t)}$ stops changing.

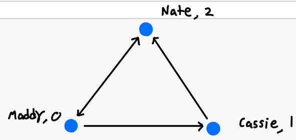
We could have chosen any eigenvalue but the Perron-Frobenius theorem tells us that we have nice properties when we choose the largest $|\lambda|$.

Code for using the power iteration method to compute the eigenvector associated with the largest |eigenvalue|

In [1]:
```
import numpy as np
```

In [2]:
```
v = np.array([1, 1, 1])   # arbitrary vector
A = np.array([[0, 1, 1], [0, 0, 1], [1, 0, 0]])    # adjacency matrix

print(v)
print(A)
```




```
[1 1 1]
[[0 1 1]
 [0 0 1]
 [1 0 0]]
```

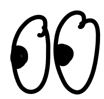
In [3]:
```
print(v)
for i in range(100):
    mult = np.matmul(A,v)
    v = mult / np.linalg.norm(mult)
    print(v)
```

```
[1 1 1]
[0.81649658 0.40824829 0.40824829]
[0.66666667 0.33333333 0.66666667]
[0.72760688 0.48507125 0.48507125]
[0.74278135 0.37139068 0.55708601]
[0.70710678 0.42426407 0.56568542]
[0.73786479 0.42163702 0.52704628]
[0.7228974  0.40160966 0.56225353]
[0.72494651 0.42288547 0.54370988]
[0.7295372  0.41036468 0.5471529 ]
[0.72413793 0.4137931  0.55172414]
[0.72757958 0.41575976 0.54568469]
[0.72646842 0.41231991 0.54975988]
[0.72610523 0.41491728 0.54828354]
[0.7269493  0.41380191 0.54800793]
[0.72626225 0.41380058 0.54891914]
[0.72658644 0.41428174 0.54812661]
[0.72655796 0.41380122 0.5485272 ]
[0.72644709 0.41407484 0.54846755]
[0.72657024 0.41400795 0.5483549 ]
    ⋮
```

This will eventually stop changing (or converge)

③ There are many other centrality measures

- PageRank: what Google used to rank websites
- motif centrality: group last year ESMI explored this
- Betweenness centrality
- closeness centrality
- Kats centrality
- ⋮



Potential Projects you can do:

- Analyze an interesing network using centrality. As we've seen, networks are every where: movies, food webs, maps, social media,... Friday were going to look at "Networks in the wild" where we will look at what networks are readily available or you can create on your own.

- Explore new centrality measures. Learn about exciting centrality measures and propose a new one.

Assignments for today:

- no worksheet for today
- finish the worksheet I gave out yesterday by Friday
- read centrality wikipedia page (linked on the website for today)
- If you have time, read up to page 26 in the textbook and do those exercises

Look here from a sample project from last year

Determining Importance of Species in Food Web Networks Through Motif-Based Centrality

Jayson Baptiste, Angelica Estevez Abad and Ilya Amburg

Cornell University

Introduction

For ecologists, networks are an extremely useful tool for representing trophic interactions between organisms. For example, in a defined ecosystem, networks can be conveniently analyzed by determining centrality which is a way to measure a node's importance. However, some forms of centrality, such as degree centrality, and pagerank centrality all have different shortcomings. Degree centrality only focuses on neighboring nodes, which in a food network means possibly neglecting how well connected a node's neighbors are. PageRank centrality ignores spammers, which are important because they serve as consumers/food for many other organisms and should not be discounted. Therefore, in this project we attempted to analyze food networks using motif centrality which counts the presence of motifs, or smaller subgraphs of nodes that represent specific patterns of interaction between species. This way, a species importance to its ecosystem is based on how many relationships would be affected if it were removed.

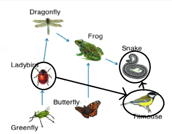


Figure: Example of In-Out Wedge

Motif Centrality

Motif Centrality

- node is important if it has both incoming and outgoing connections

$$x_i = \Sigma_{j=1}^{n} A_{ij} * \Sigma_{j=1}^{n} A_{ji} \quad (1)$$

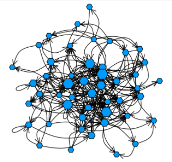


Figure: Motif Centrality Graph of St. Marks National Wildlife Refuge, Florida (node size represents centrality)

Established Centrality Measures

Pagerank Centrality

- node is important if highly linked
- node is important if linked to other highly linked nodes that don't overlink

$$A = \alpha P + \frac{1}{n}(1 - \alpha)1 * 1^T \quad (2)$$

$$Ax = x \quad (3)$$

- A is the adjacency matrix, connections are directed
- α is a number between 0 and 1
- P is a normalized adjacency matrix who's columns sum to 1

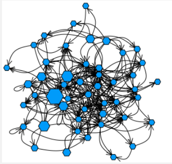


Figure: Pagerank Graph of St. Marks National Wildlife Refuge, Florida (node size represents centrality)

Degree Centrality

- node is important if there are many edges entering/leaving

$$x_i = \Sigma_{j=1}^{n} A_{ij} + \Sigma_{j=1}^{n} A_{ji} \quad (4)$$

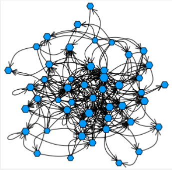


Figure: Degree Graph of St. Marks National Wildlife Refuge, Florida (node size represents centrality)

Methods and Results

Methods

- We implemented our approach using Julia
- Parsed the raw data into adjacency matrices
- created functions to implement measures of centrality
- Examined the 10 highest species for each centrality:

	Degree	PageRank	Motif
1	Benthic C.	Phytoplankton	Benthic C.
2	Shrimps	Detrius	Crabs
3	Crabs	M&M Zooplankton	Shrimps
4	Benthopelagic C.	Shrimps	Benthopelagic C.
5	S.D fishes	Suprabenthos	S. D. Fishes
6	Sharks	Macrozooplankton	Flatfishes
7	D. Piscivores	Benthic C.	Demersal fishes
8	Flatfishes	S. B. Fishes	R. Shrimp
9	Demersal fishes	B. Invertebrates FDS	Juvenile hake
10	Suprabenthos	Bivalves	B. Invertebrates

Table: Centrality Rankings FW-005

- Computed the Spearman rank correlation coefficient to determine the similarity among the rankings:

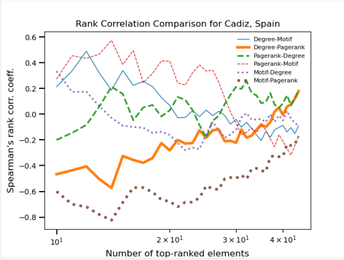


Figure: Rank Correlation for Cadiz, Spain(FW-005)

Conclusion

The goal of this research project was to determine if motif centrality is a viable method of classifying the importance of nodes in a directed network. Looking at the St.Marks National Wild refuge ecosystem, motif centrality provides a more uniform distribution of node weights. By comparison the node size distributions for degree and pagerank are both skewed towards small values with pagerank having sharp peaks at several nodes. When observing their rankings we noticed that the ten highest ranked species for the different centralities also feature some overlap but are still noticeably different, indicating motif centrality has viability compared to established centrality measures.

References

1. Raul Ortega, Miguel A. Fortuna, Jordi Bascompte, WWW.web-of-life.es/
2. FW-005:Torres MA., Coll M., Heymans JJ., Christensen V,. Sobrino I. (2013) Food-web structure of and fishing impacts on the Gulf of Cadiz ecosystem (South-western Spain). Ecological Modelling 265 (2013) 26–44
3. FW-007:Christian, RR., Luczkovich JJ. (1999) Organizing and understanding a winter's seagrass foodweb network through effective trophic levels
4. Austin R. Benson, . "Three hypergraph eigenvector centralities." (2019).

Support Information

Research supported by Engineering Summer Math Institute (ESMI)—Cornell University Engineering Success (CUES), supported by the Fund II Foundation Endowment, Cornell Engineering.

Contact Information

- Angelica Estevez Abad, Email: ae334@cornell.edu
- Jayson Baptiste, Email: jcb482@cornell.edu