9. Graph Metrics – A Measure of Clustering and Modularity

Bookmark this page

Exercises due Oct 20, 2021 17:29 IST Completed

Graph Metrics – A Measure of Clustering and Modularity

Start of transcript. Skip to the end.

Prof Uhler: So good-- so this is about distances between them.

And we'll end by looking at clustering coefficients

and how to see whether more similar nodes cluster

within each other.

So this also, we discussed a little bit

in the first three videos.

We, in particular, discussed how we

0:00 / 0:00 2.0x

Video
Download video file

Transcripts
Download SubRip (.srt) file
Download Text (.txt) file

Triangle Density and Clustering Coefficient

The **triangle density** of a graph is the ratio of number of triangles in the graph to the number of possible triangles:

$$\text{triangle density} \triangleq \frac{\#\text{ of triangles}}{\binom{n}{3}}$$

Triangle density is not an appropriate metric to measure clustering for several reasons. First, it does not take into account that the graph could have several connected components, in which case the denominator might be much larger than the numerator. Second, even in the case of a connected graph any three nodes need not be present in the same cluster (the shortest path lengths connecting them may be much larger).

A better metric for clustering is the **clustering coefficient** , denoted C, which measures the ratio of triangles in the network to the number of connected triples:

$$C \quad = \frac{\#\text{ of closed triplets}}{\#\text{ of closed and open triplets}} \quad = \frac{3 \cdot \#\text{ of triangles}}{\#\text{ of connected triples}}.$$

where an open triplet is three nodes connected by two edges, and a closed triplet is three nodes connected by three edges. This can be written in terms of the adjacency matrix as

$$C \quad = \frac{\sum_{i,j,k} A_{ij} A_{jk} A_{ki}}{\sum_i k_i \left(k_i - 1\right)},$$

where $k_i = \sum_j A_{ij}$ is the degree of node i.

To understand this formula, first consider the numerator. Note that $\sum_{i,j,k} A_{ij} A_{jk} A_{ki} = \sum_i \left[A^3\right]_{ii} = \mathbf{tr}\, A^3$, that is, the trace of A^3. We know that $\left[A^3\right]_{ii}$ is equal to the number of walks of length 3 from node i to itself, which will be two if it is part of a closed triplet (there are two paths around the triplet) and zero otherwise. So the sum of the diagonal elements of A^3 is exactly twice the number of closed triplets (and six times the number of triangles, as each node in the triangle is counted once).

As for the denominator, let us examine how the degree of a node informs the number of connected triplets. If a node has degree zero, then it can't be part of a triplet, and the same is true for degree one. For a node of degree two, it must be part of one triplet (which may be closed or open). For degree three, the node is part of three triplets. We conclude that for a node of degree k, the node is part of $\binom{k}{2} = k(k-1)/2$ connected triplets. Therefore the total number of connected triplets is the sum of this formula for all nodes: $\sum_i k_i (k_i - 1)/2$.

One can also define the same node-wise. For node i, the local clustering coefficient C_i is defined as

$$\begin{aligned} C_i &= \frac{\#\text{ of triangles at node } i}{\#\text{ of connected triples centered at node } i} \\ &= \frac{\sum_{j,k} A_{ij} A_{jk} A_{ki}}{k_i (k_i - 1)} \end{aligned}$$

Clustering Coefficient of an Almost Complete Graph

1/1 point (graded)

A *complete graph* is an undirected graph on n nodes such that every node is connected to every other node. Say you remove an edge from a complete graph on n nodes. What is the new clustering coefficient? Assume that $n \geq 3$.

(n^2-n-6)/(n^2-n-4)

✔ **Answer:** 3*(n*(n-1)*(n-2)/6 - (n-2))/(n*(n-1)*(n-2)/2 - 2*(n-2))

Solution:

In a complete graph we have $\binom{n}{3}$ triangles and $3 \cdot \binom{n}{3}$ connected triples so that the clustering coefficient is equal to 1. Now, removing an edge results in the loss of $n-2$ triangles. In addition, the number of connected triples is $n\binom{n-1}{2}$ in the complete graph, and $(n-2)\binom{n-1}{2} + 2\binom{n-2}{2}$ in the modified graph. Therefore, removing an edge results in the loss of $2(n-2)$ connected triples.

The new clustering coefficient is

$$\frac{3 \cdot \left(\binom{n}{3} - (n-2)\right)}{3 \cdot \binom{n}{3} - 2(n-2)}.$$

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

Modularity

The **modularity** of an undirected graph with node types $t_i, i = 1, \ldots, n$ is defined as

$$\frac{1}{2m} \sum_{i,j} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(t_i, t_j),$$

where A is the adjacency matrix, m is the number of edges, and $\delta(t_i, t_j) = 1$ if $t_i = t_j$ and is equal to 0 if $t_i \neq t_j$. A few things to note:

- The definition of modularity is a little abstract, but the main thing to note here is that the expected number of edges between a node i with degree k_i and a node j with degree k_j in a random graph with m edges according to the *configuration model* (to be explained in a later lecture) is equal to $\frac{k_i k_j}{2m}$.

- For a given pair of nodes i, j, a positive value of $\left(A_{ij} - \frac{k_i k_j}{2m}\right) \delta(t_i, t_j)$ indicates that nodes i, j have an affinity that is more than the expected affinity that they would otherwise have in a truly random graph obtained according to the configuration model with given node types and m edges.
- Likewise, a negative value of $\left(A_{ij} - \frac{k_i k_j}{2m}\right) \delta(t_i, t_j)$ indicates that nodes i, j have lesser than expected affinity when compared to a random graph with the same characteristics.

Modularity of a Small Graph

1/1 point (graded)

Consider the following adjacency matrix of an undirected graph:

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

Raw matrix

Python:

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

Mathematica:

```
{{0, 1, 0, 1, 0, 1, 0, 0, 0, 0},
{1, 0, 1, 1, 1, 0, 0, 0, 0, 0},
{0, 1, 0, 0, 0, 0, 0, 0, 0, 0},
{1, 1, 0, 0, 0, 1, 0, 1, 1, 0},
{0, 1, 0, 0, 0, 0, 0, 0, 1, 1},
{1, 0, 0, 1, 0, 0, 0, 0, 0, 0},
{0, 0, 0, 0, 0, 0, 0, 0, 0, 1},
{0, 0, 0, 1, 0, 0, 0, 0, 0, 0},
{0, 0, 0, 1, 1, 0, 0, 0, 0, 1},
{0, 0, 0, 0, 1, 0, 1, 0, 1, 0}}
```

Hide

Assume that the nodes are numbered $0, \ldots, 9$. Let there be two node types: nodes $0, 2, 4, 6, 8$ are of type 1 and nodes $1, 3, 5, 7, 9$ are of type 2. Compute the modularity of the graph. Provide an answer accurate to at least three **significant figures** (graded to 1% tolerance).

-0.20414201183431932

✔ **Answer:** -0.2041

Solution:

Python:

```
def modularity_partition(A, part):
        m = A.sum()/2
        ks = A.sum(axis=0)
        return ( A[part].T[part].T - ks[part][:,None]*ks[part][None,:]/(2*m) ).sum()/(2*m)

def modularity(A, parts):
        return sum([ modularity_partition(A, p) for p in parts ])

modularity(mat, [[0,2,4,6,8], [1,3,5,7,9]])
```

Mathematica:

```
Needs["GraphUtilities`"]
CommunityModularity[A, {{1, 3, 5, 7, 9}, {2, 4, 6, 8, 10}}]
```

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Module 3: Network Analysis:Graph Basics / 9. Graph Metrics – A Measure of Clustering and Modularity

Add a Post

Show all posts by recent activity

- how to enter n choose x please ? — as question title, thanks — 3
- Modularity of small Graph - Code — Hey there, here's some code to make your life a little easier: The **networkx library** offers a lot of useful functionalities when it co... — 5
- Could someone clarify for me the second reason for why triangular density is not a good metric? — I am not able to get this sentence clearly: *Second, even in the case of a connected graph any three nodes need not be present in t... — 2
- Clustering Coefficient of an Almost Complete Graph — Staff: Could you please take a look at my answer to this question? The derived expression gives correct answers for clustering coeffi... — 2
- [hint] Clustering Coefficient of an Almost Complete Graph — The metric you are looking for in networkx module is transitivity (to verify your formula) — 3
- question about definition of degree of a node — 3
- STAFF: clustering coefficient on complete graph — Hi, could you please check the answer provided for that question. I've checked my result for different values of n and it seems right. ... — 1
- Clustering coefficient formula correction — 1
- Clustering Coefficient of an Almost Complete Graph — Hello, I think my answer to the question on Clustering Coefficient of an Almost Complete Graph is correct, please check. — 1
- Modularity of a Small Graph — hi! What is the meaning about a node is type one or type two? thanks — 2

< Previous Next >

edX

[About]
[Affiliates]
[edX for Business]
[Open edX]
[Careers]
[News]

Legal

[Terms of Service & Honor Code]
[Privacy Policy]
[Accessibility Policy]
[Trademark Policy]
[Sitemap]

Connect

[Blog]
[Contact Us]
[Help Center]
[Media Kit]
[Donate]

GET IT ON Google play
Download on the App Store