WIKIPEDIA

Clustering coefficient

In graph theory, a **clustering coefficient** is a measure of the degree to which nodes in a graph tend to cluster together. Evidence suggests that in most real-world networks, and in particular social networks, nodes tend to create tightly knit groups characterised by a relatively high density of ties; this likelihood tends to be greater than the average probability of a tie randomly established between two nodes (Holland and Leinhardt, 1971;[1] Watts and Strogatz, 1998[2]).

Two versions of this measure exist: the global and the local. The global version was designed to give an overall indication of the clustering in the network, whereas the local gives an indication of the embeddedness of single nodes.

Contents

Local clustering coefficient

The **local clustering coefficient** of a vertex (node) in a graph quantifies how close its neighbours are to being a clique (complete graph). Duncan J. Watts and Steven Strogatz introduced the measure in 1998 to determine whether a graph is a small-world network.

A graph $G = (V, E)$ formally consists of a set of vertices V and a set of edges E between them. An edge e_{ij} connects vertex v_i with vertex v_j.

The neighbourhood N_i for a vertex v_i is defined as its immediately connected neighbours as follows:

$$N_i = \{v_j : e_{ij} \in E \lor e_{ji} \in E\}.$$

We define k_i as the number of vertices, $|N_i|$, in the neighbourhood, N_i, of a vertex.

The local clustering coefficient C_i for a vertex v_i is then given by a proportion of the number of links between the vertices within its neighbourhood divided by the number of links that could possibly exist between them. For a directed graph, e_{ij} is distinct from e_{ji}, and therefore for each neighbourhood N_i there are $k_i(k_i - 1)$ links that could exist among the vertices within the neighbourhood (k_i is the number of neighbours of a vertex). Thus, the **local clustering coefficient for directed graphs** is given as [2]

$$C_i = \frac{|\{e_{jk} : v_j, v_k \in N_i, e_{jk} \in E\}|}{k_i(k_i - 1)}.$$

An undirected graph has the property that e_{ij} and e_{ji} are considered identical. Therefore, if a vertex v_i has k_i neighbours, $\dfrac{k_i(k_i - 1)}{2}$ edges could exist among the vertices within the neighbourhood. Thus, the **local clustering coefficient for undirected graphs** can be defined as

$$C_i = \frac{2|\{e_{jk} : v_j, v_k \in N_i, e_{jk} \in E\}|}{k_i(k_i - 1)}.$$

Let $\lambda_G(v)$ be the number of triangles on $v \in V(G)$ for undirected graph G. That is, $\lambda_G(v)$ is the number of subgraphs of G with 3 edges and 3 vertices, one of which is v. Let $\tau_G(v)$ be the number of triples on $v \in G$. That is, $\tau_G(v)$ is the number of subgraphs (not necessarily induced) with 2 edges and 3 vertices, one of which is v and such that v is incident to both edges. Then we can also define the clustering coefficient as

$$C_i = \frac{\lambda_G(v)}{\tau_G(v)}.$$

It is simple to show that the two preceding definitions are the same, since

$$\tau_G(v) = C(k_i, 2) = \frac{1}{2}k_i(k_i - 1).$$

These measures are 1 if every neighbour connected to v_i is also connected to every other vertex within the neighbourhood, and 0 if no vertex that is connected to v_i connects to any other vertex that is connected to v_i.

Since any graph is fully specified by its adjacency matrix A, the local clustering coefficient for a simple undirected graph can be expressed in terms of A as:[3]

$$C_i = \frac{1}{k_i(k_i - 1)} \sum_{j,k} A_{ij} A_{jk} A_{ki}$$
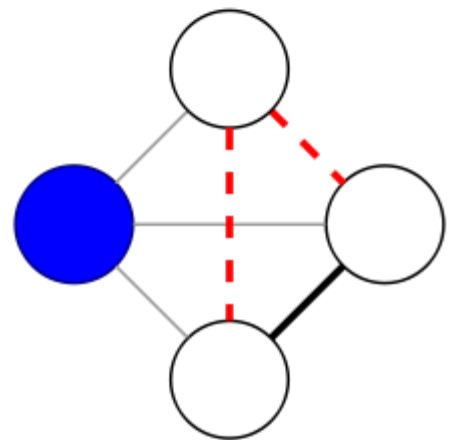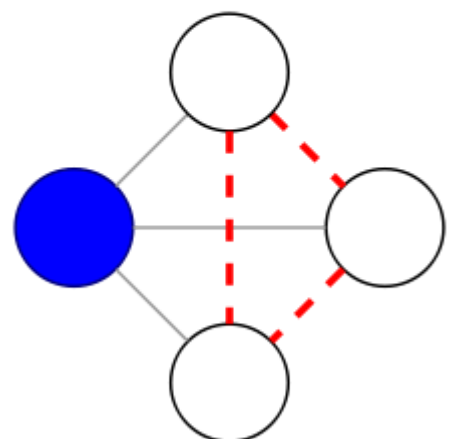
where:

$$k_i = \sum_j A_{ij}$$

and C_i=0 when k_i is zero or one. In the above expression, the numerator counts twice the number of complete triangles that vertex i is involved in. In the denominator, k_i^2 counts the number of edge pairs that vertex i is involved in plus the
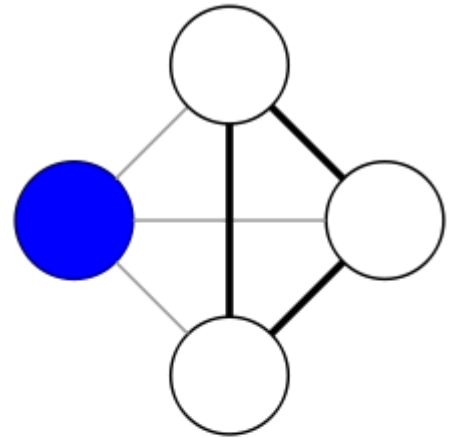
C = 1

C = 1/3

C = 0

Example local clustering coefficient on an undirected graph. The local clustering coefficient of the blue node is computed as the proportion of connections among its

number of single edges traversed twice. k_i is the number of edges connected to vertex i, and subtracting k_i then removes the latter, leaving only a set of edge pairs that could conceivably be connected into triangles. For every such edge pair, there will be another edge pair which could form the same triangle, so the denominator counts twice the number of conceivable triangles that vertex i could be involved in.

neighbours which are actually realised compared with the number of all possible connections. In the figure, the blue node has three neighbours, which can have a maximum of 3 connections among them. In the top part of the figure all three possible connections are realised (thick black segments), giving a local clustering coefficient of 1. In the middle part of the figure only one connection is realised (thick black line) and 2 connections are missing (dotted red lines), giving a local cluster coefficient of 1/3. Finally, none of the possible connections among the neighbours of the blue node are realised, producing a local clustering coefficient value of 0.

Global clustering coefficient

The **global clustering coefficient** is based on triplets of nodes. A triplet is three nodes that are connected by either two (open triplet) or three (closed triplet) undirected ties. A triangle graph therefore includes three closed triplets, one centered on each of the nodes (n.b. this means the three triplets in a triangle come from overlapping selections of nodes). The global clustering coefficient is the number of closed triplets (or 3 x triangles) over the total number of triplets (both open and closed). The first attempt to measure it was made by Luce and Perry (1949).[4] This measure gives an indication of the clustering in the whole network (global), and can be applied to both undirected and directed networks (often called transitivity, see Wasserman and Faust, 1994, page 243[5]).

The global clustering coefficient is defined as:

$$C = \frac{\text{number of closed triplets}}{\text{number of all triplets (open and closed)}}.$$

The number of closed triplets has also been referred to as 3 × triangles in the literature, so:

$$C = \frac{3 \times \text{number of triangles}}{\text{number of all triplets}}.$$

A generalisation to weighted networks was proposed by Opsahl and Panzarasa (2009),[6] and a redefinition to two-mode networks (both binary and weighted) by Opsahl (2009).[7]

Since any graph is fully specified by its adjacency matrix A, the global clustering coefficient for an undirected graph can be expressed in terms of A as:

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

where:

$$k_i = \sum_j A_{ij}$$

and C=0 when the denominator is zero.

Network average clustering coefficient

As an alternative to the global clustering coefficient, the overall level of clustering in a network is measured by Watts and Strogatz[2] as the average of the local clustering coefficients of all the vertices n :[8]

$$\bar{C} = \frac{1}{n} \sum_{i=1}^{n} C_i.$$

It is worth noting that this metric places more weight on the low degree nodes, while the transitivity ratio places more weight on the high degree nodes.

A graph is considered small-world, if the graph has a small mean-shortest path length that scales with the natural log of the number of nodes in the network, $\ln N$.[9] For example, a random graph is small-world, while a lattice is not, and scale-free graphs are often considered ultra-small world, as their mean-shortest path length scales with $\ln \ln N$.

A generalisation to weighted networks was proposed by Barrat et al. (2004),[10] and a redefinition to bipartite graphs (also called two-mode networks) by Latapy et al. (2008)[11] and Opsahl (2009).[7]

Alternative generalisations to weighted and directed graphs have been provided by Fagiolo (2007)[12] and Clemente and Grassi (2018).[13]

This formula is not, by default, defined for graphs with isolated vertices; see Kaiser (2008)[14] and Barmpoutis et al.[15] The networks with the largest possible average clustering coefficient are found to have a modular structure, and at the same time, they have the smallest possible average distance among the different nodes.[15]

Percolation of clustered networks

For studying the robustness of clustered networks a percolation approach is developed.[16][17][18] The robustness with respect to localized attacks has been studied using localized percolation.[19] Wave localization in clustered complex networks has also been studied.[20]

See also

- Directed graph
- Graph theory
- Network theory
- Network science
- Percolation theory
- Scale free network
- Small world

References

1. P. W. Holland & S. Leinhardt (1971). "Transitivity in structural models of small groups". *Comparative Group Studies*. **2** (2): 107–124. doi:10.1177/104649647100200201 (https://doi.org/10.1177%2F104649647100200201). S2CID 145544488 (https://api.semanticscholar.org/CorpusID:145544488).

2. D. J. Watts & Steven Strogatz (June 1998). "Collective dynamics of 'small-world' networks". *Nature*. **393** (6684): 440–442. Bibcode:1998Natur.393..440W (https://ui.adsabs.harvard.edu/abs/1998Natur.393..440W). doi:10.1038/30918 (https://doi.org/10.1038%2F30918). PMID 9623998 (https://pubmed.ncbi.nlm.nih.gov/9623998). S2CID 4429113 (https://api.semanticscholar.org/CorpusID:4429113).

3. Wang, Yu; Ghumare, Eshwar; Vandenberghe, Rik; Dupont, Patrick (2017). "Comparison of Different Generalizations of Clustering Coefficient and Local Efficiency for Weighted Undirected Graphs" (https://www.mitpressjournals.org). *Neural Computation*. **29** (2): 313–331. doi:10.1162/NECO_a_00914 (https://doi.org/10.1162%2FNECO_a_00914). PMID 27870616 (https://pubmed.ncbi.nlm.nih.gov/27870616). S2CID 11000115 (https://api.semanticscholar.org/CorpusID:11000115). Retrieved August 8, 2020.

4. R. D. Luce & A. D. Perry (1949). "A method of matrix analysis of group structure". *Psychometrika*. **14** (1): 95–116. doi:10.1007/BF02289146 (https://doi.org/10.1007%2FBF02289146). hdl:10.1007/BF02289146 (https://hdl.handle.net/10.1007%2FBF02289146). PMID 18152948 (https://pubmed.ncbi.nlm.nih.gov/18152948). S2CID 16186758 (https://api.semanticscholar.org/CorpusID:16186758).

5. Stanley Wasserman, Katherine Faust, 1994. *Social Network Analysis: Methods and Applications.* Cambridge: Cambridge University Press.

6. Tore Opsahl & Pietro Panzarasa (2009). "Clustering in Weighted Networks" (http://toreopsahl.com/2009/04/03/article-clustering-in-weighted-networks/). *Social Networks*. **31** (2): 155–163. doi:10.1016/j.socnet.2009.02.002 (https://doi.org/10.1016%2Fj.socnet.2009.02.002).

7. Tore Opsahl (2009). "Clustering in Two-mode Networks" (http://toreopsahl.com/2009/09/11/clustering-in-two-mode-networks/). *Conference and Workshop on Two-Mode Social Analysis (Sept 30-Oct 2, 2009)*.

8. Kemper, Andreas (2009). *Valuation of Network Effects in Software Markets: A Complex Networks Approach* (https://books.google.com/books?id=isdza0IiXbUC&pg=PA142). Springer. p. 142. ISBN 9783790823660.

9. *Network Science by Albert-László Barabási* (http://networksciencebook.com/4#ultra-small).

10. Barrat, A.; Barthelemy, M.; Pastor-Satorras, R.; Vespignani, A. (2004). "The architecture of complex weighted networks" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC374315). *Proceedings of the National Academy of Sciences*. **101** (11): 3747–3752. arXiv:cond-mat/0311416 (https://arxiv.org/abs/cond-mat/0311416). Bibcode:2004PNAS..101.3747B (https://ui.adsabs.harvard.edu/abs/2004PNAS..101.3747B). doi:10.1073/pnas.0400087101 (https://doi.org/10.1073%2Fpnas.0400087101). PMC 374315 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC374315). PMID 15007165 (https://pubmed.ncbi.nlm.nih.gov/15007165).

11. Latapy, M.; Magnien, C.; Del Vecchio, N. (2008). "Basic Notions for the Analysis of Large Two-mode Networks". *Social Networks*. **30** (1): 31–48. doi:10.1016/j.socnet.2007.04.006 (https://doi.org/10.1016%2Fj.socnet.2007.04.006).

12. Fagiolo, G. (2007). "Clustering in complex directed networks". *Physical Review E*. **76** (2 Pt 2): 026107. CiteSeerX 10.1.1.262.1006 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.262.1006). doi:10.1103/PhysRevE.76.026107 (https://doi.org/10.1103%2FPhysRevE.76.026107). PMID 17930104 (https://pubmed.ncbi.nlm.nih.gov/17930104). S2CID 2317676 (https://api.semanticscholar.org/CorpusID:2317676).

13. Clemente, G.P.; Grassi, R. (2018). "Directed clustering in weighted networks: A new perspective". *Chaos, Solitons & Fractals*. **107**: 26–38. arXiv:1706.07322 (https://arxiv.org/abs/1706.07322). Bibcode:2018CSF...107...26C (https://ui.adsabs.harvard.edu/abs/2018CSF...107...26C). doi:10.1016/j.chaos.2017.12.007 (https://doi.org/10.1016%2Fj.chaos.2017.12.007). S2CID 21919524 (https://api.semanticscholar.org/CorpusID:21919524).

14. Kaiser, Marcus (2008). "Mean clustering coefficients: the role of isolated nodes and leafs on clustering measures for small-world networks". *New Journal of Physics*. **10** (8): 083042. arXiv:0802.2512 (https://arxiv.org/abs/0802.2512). Bibcode:2008NJPh...10h3042K (https://ui.adsabs.harvard.edu/abs/2008NJPh...10h3042K). doi:10.1088/1367-2630/10/8/083042 (https://doi.org/10.1088%2F1367-2630%2F10%2F8%2F083042). S2CID 16480565 (https://api.semanticscholar.org/CorpusID:16480565).

15. Barmpoutis, D.; Murray, R. M. (2010). "Networks with the Smallest Average Distance and the Largest Average Clustering". arXiv:1007.4031 (https://arxiv.org/abs/1007.4031) [q-bio.MN (https://arxiv.org/archive/q-bio.MN)].

16. M. E. J. Newman (2009). "Random Graphs with Clustering". *Phys. Rev. Lett.* **103** (5): 058701. arXiv:0903.4009 (https://arxiv.org/abs/0903.4009). doi:10.1103/PhysRevLett.103.058701 (https://doi.org/10.1103%2FPhysRevLett.103.058701). PMID 19792540 (https://pubmed.ncbi.nlm.nih.gov/19792540). S2CID 28214709 (https://api.semanticscholar.org/CorpusID:28214709).

17. A. Hackett; S. Melnik & J. P. Gleeson (2011). "Cascades on a class of clustered random networks". *Phys. Rev. E*. **83** (5 Pt 2): 056107. arXiv:1012.3651 (https://arxiv.org/abs/1012.3651). doi:10.1103/PhysRevE.83.056107 (https://doi.org/10.1103%2FPhysRevE.83.056107). PMID 21728605 (https://pubmed.ncbi.nlm.nih.gov/21728605). S2CID 18071422 (https://api.semanticscholar.org/CorpusID:18071422).

18. S. Shao; X. Huang; H.E. Stanley; S. Havlin (2014). "Robustness of a partially interdependent network formed of clustered networks". *Phys. Rev. E*. **89** (3): 032812. arXiv:1308.0034 (https://arxiv.org/abs/1308.0034). doi:10.1103/PhysRevE.89.032812 (https://doi.org/10.1103%2FPhysRevE.89.032812). PMID 24730904 (https://pubmed.ncbi.nlm.nih.gov/24730904). S2CID 14685627 (https://api.semanticscholar.org/CorpusID:14685627).

19. Fan Wang; Ruijin Du; Lixin Tian; Shuai Shao; H Eugene Stanley; Shlomo Havlin (2019). "Localized attack on networks with clustering Huifang Hao". *New Journal of Physics*. **21**: 013014. doi:10.1088/1367-2630/aaf773 (https://doi.org/10.1088%2F1367-2630%2Faaf773).

20. L. Jahnke; J.W. Kantelhardt; R. Berkovits; S. Havlin Phys (2008). "Wave Localization in Complex Networks with High Clustering". *Phys. Rev. Lett.* **101** (17): 175702. doi:10.1103/PhysRevLett.101.175702 (https://doi.org/10.1103%2FPhysRevLett.101.175702). PMID 18999764 (https://pubmed.ncbi.nlm.nih.gov/18999764).

External links

- Media related to Clustering coefficient at Wikimedia Commons

This page was last edited on 25 August 2021, at 08:34 (UTC).