Part I
Normalized clustering coefficients

1 Introduction

1.1 Current measures for clustering

Local clustering coefficient [1]:

$$c_i = \begin{cases} \dfrac{2T_i}{k_i(k_i-1)} & \text{if } k_i > 1 \\ 0 & \text{if } k_i \leq 1, \end{cases} \tag{1}$$

where T_i is the number of triangles passing through vertex i and k_i is its degree.

Average Watts-Strogatz clustering coefficient:

$$\bar{C} = \frac{1}{N}\sum_i c_i. \tag{2}$$

Obs: This average is taken over all the nodes in the network. Some authors use, instead, the only the nodes with degree greater than 1. It is important to be aware of that when interpreting results of other authors.

Degree-dependent clustering coefficient [2]:

$$\bar{c}(k) = \frac{1}{N_k}\sum_{i\in Y(k)} c_i = \frac{1}{k(k-1)N_k}\sum_{i\in Y(k)} 2T_i, \tag{3}$$

where N_k is the number of nodes with degree k and $Y(k)$ is the set of such nodes.

The quantities $\bar{C}$ and $\bar{c}(k)$ are related by

$$\bar{c} = \sum_k p(k)\bar{c}(k), \tag{4}$$

$p(k)$ being the degree distribution.

A different global transitivity measure C was introduced by Newman in [3], and is defined as

$$C = \frac{2T}{\sum_i k_i(k_i-1)}, \tag{5}$$

where $T = \sum_i T_i$ is the number of triangles in the network.

The two global clustering coefficients have similar definitions and, in fact, for some networks both have similar values. In particular, for an uncorrelated network, $\bar{C}$, $\bar{c}(k)$ and C are identical and its value can be computed as [4]

$$C_{\text{unc}} = \frac{1}{N} \frac{\left[\langle k^2 \rangle - \langle k \rangle\right]^2}{\langle k \rangle^3}. \tag{6}$$

Nevertheless, these coefficients do differ significantly in some cases, as can be seen in [5, 6].

1.2 Normalized clustering coefficient

We define the normalized clustering coefficients

$$C_{\text{norm}} = \frac{C - C_{\text{rand}}}{C_{\text{max}} - C_{\text{rand}}} \qquad \text{and} \qquad \bar{c}_{\text{norm}} = \frac{\bar{c} - \bar{c}_{\text{rand}}}{\bar{c}_{\text{max}} - \bar{c}_{\text{rand}}} \tag{7}$$

C_{rand} and $\bar{c}_{\text{rand}}$ represent the average of the clustering coefficient computed for an ensemble of random networks with the same degree sequence as the original network, and C_{max} and $\bar{c}_{\text{max}}$ correspond to the maximum clustering values that can be achieved over the networks in the ensemble. In our work, we will consider the ensamble of all networks having the same degree sequence as the original network.

1.3 Maximum clustering coefficient

The definitions given by equation 6 require the knowledge of the greatest value for the clustering coefficient that can be achieved by a graph having a specified degree sequence. In this section we will address how to compute that value.

1.3.1 Approximation of C_{max}

Havel-Hakimi algorithm:

This is a known algorithm that is very useful because it allows to determine whether a given degree sequence is graphic or not [7]. It is based on the Havel-Hakimi theorem, which states that the degree sequence $S = [k_1, k_2, ..., k_N]$, where $k_i \geq k_j$, $\forall i \leq j$, is graphic if and only if the sequence $S' = [k_2 - 1, k_3 - 1, ..., k_{k_1+1} - 1, k_{k_1+2}, ..., k_N]$ is also graphic. If the given list S is graphic, then the theorem will be applied at most $N - 1$ times setting in each further step $S := S'$. Note that it can be necessary to sort this list again. This process ends when the whole list S' consists of zeros. In each step of the algorithm one constructs the edges of a graph with vertices $v_1, ..., v_N$, i.e. if it is possible to reduce the list S to S', then we add edges $\{v_1, v_2\}, \{v_1, v_3\}, \{v_1, v_{d_1+1}\}$,. When the list S cannot be reduced to a list S' of non-negative integers in any step of this approach, the theorem proves that the list S from the beginning is not graphic.

The advantage of this algorithm is that it always converges when the original sequence is graphic and that the resulting graph es quite clusterized. The disadvantage is that in general it doesn't give the most clusterized graph. For

example, given the degree sequence $S_1 = [3, 3, 2, 2, 2, 1, 1]$, the result of the algorithm is the graph in left side of figure 1, whilst the most clustered graph for that degree sequence is the graph in the right part of the figure.

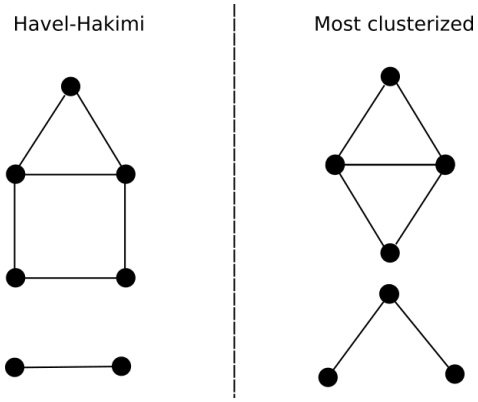

Figure 1: Graphs obtained from the degree sequence $S_1 = [3, 3, 2, 2, 2, 1, 1]$

Greedy algorithm:

This algorithm is similar as the Havel-Hakimi algorithm, with the difference that at each step the list is not sorted. This algorithm gives very good results for most of the real-world networks we studied (the resulting graph is more clusterized than the graph using the Havel-Hakimi algorithm). In particular, in the example in figure 1, it finds the most clusterized graph. Also, comparing with Monte Carlo simulations, it seems that this algorithm finds graphs with a clustering coefficient very close to the maximum.

The main disadvantage of this algorithm is that it doesn't always converge. One counterexample is the degree sequence $S_2 = [3, 3, 2, 2, 2, 2]$. After adding the edges $\{v_1, v_2\}, \{v_1, v_3\}, \{v_1, v_4\}, \{v_2, v_3\}, \{v_2, v_4\}, \{v_5, v_6\}$, the algorithm is stuck, as there is no possible pair of nodes where to put the last edge. To overcome this flaw, we perform the following modification. Whenever the algorithm get stuck, lets say after adding the edge $\{v_i, v_j\}$, we remove this edge and try to connect node v_i with v_{j+1}. If the algorithm stuck after adding the edge $\{v_i, v_N\}$, we try with $\{v_{i+1}, v_j\}$. This way, the algorithm always converges.

In most cases, this algorithm seems to converge to graphs very close to the most clustered graph. But if we see the example S_2, we can show that the result of the algorithm is the graph in the right side of figure

References

[1] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440–442, 6 1998.

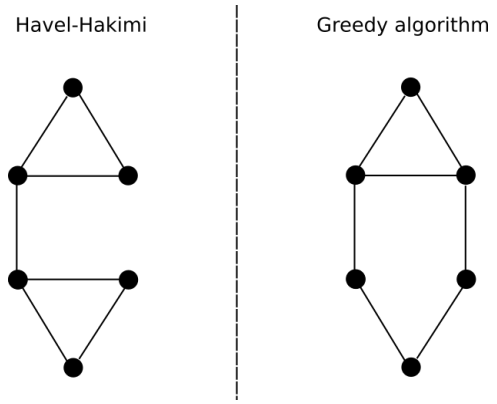

Figure 2: Graphs obtained from the degree sequence $S_1 = [3, 3, 2, 2, 2, 2]$

[2] A. Vázquez, R. Pastor-Satorras, and A. Vespignani, "Large-scale topological and dynamical properties of the Internet," *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 65, no. 6, pp. 1–12, 2002.

[3] M. E. J. Newman, "Mixing patterns in networks," *Physical Review E*, vol. 67, 2 2002.

[4] M. E. J. Newman, *Networks*. Oxford University Press, 2 ed., 2018.

[5] B. Bollobás and O. M. Riordan, "Mathematical results on scale-free random graphs," in *Handbook of Graphs and Networks*, pp. 1–34, Wiley-VCH Verlag GmbH & Co. KGaA, dec 2004.

[6] E. Estrada, *The Structure of Complex Networks*. Oxford University Press, oct 2011.

[7] H. S., "On realizability of a set of integers as degrees of the vertices of a linear graph. i," *Journal of SIAM*, vol. 10, pp. 496–506, 1962.