

# Small subgraphs: node clustering coefficient and reciprocated arcs

Francesco Coghi<sup>1, a)</sup>

*School of Mathematical Sciences, Queen Mary University of London, London E1 4NS, UK*

(Dated: 22 March 2018)

## CONTENTS

<b>I. Node clustering coefficient on undirected random regular graphs</b>	2
A. Background	2
B. Calculation of the clustering coefficient for some graphs	2
1. Erdős-Rényi random graphs	2
2. Random regular graphs	3
C. Numerical simulations to confirm the results	3
1. Erdős-Rényi random graphs	4
2. Random regular graphs	4
<b>II. Reciprocated arcs in directed regular random graphs</b>	4
A. Calculation of the expected fraction of reciprocated arcs in a directed random regular graph	5

---

<sup>a)</sup>Electronic mail: [f.coghi@qmul.ac.uk](mailto:f.coghi@qmul.ac.uk)

## I. NODE CLUSTERING COEFFICIENT ON UNDIRECTED RANDOM REGULAR GRAPHS

### A. Background

Here we focus on the presence of triangles in undirected random regular graphs  $G = (V, E)$ . There are two alternative measures to quantify the presence of triangles in a graph (cite book Vito Vincenzo).

The first measure is called **graph transitivity**  $T$  and it is defined as follows:

$$T = \frac{3(\text{number of triangles})}{(\text{number of triads})} = \frac{3n_{\Delta}}{n_{\wedge}} \quad (1)$$

where  $n_{\Delta}$  is the number of triangles and  $n_{\wedge}$  is the number of triads (trees of three nodes) in the graph.

The second measure is the **clustering coefficient** and it averages local information coming from the node clustering coefficients

$$c_i = \begin{cases} \frac{K[G_i]}{\binom{k_i}{2}} & \text{for } k_i \geq 2 \\ 0 & \text{for } k_i = 0, 1 \end{cases} \quad (2)$$

where  $K[G_i]$  represents the number of links, and  $\binom{k_i}{2}$  its theoretical maximum, in  $G_i$  the subgraph induced by the neighbours of  $i$ . From this, the graph clustering coefficient  $C$ , is obtained just by averaging over all the nodes as follows:

$$C = \langle c \rangle = \frac{1}{N} \sum_{i \in V} c_i. \quad (3)$$

$C$  and  $T$  are in principle similar measures, nevertheless on several graphs they could give quite different results. This is due to the fact that

$$T = \frac{\sum_i n_i^{\wedge} c_i}{\sum_i n_i^{\wedge}}, \quad (4)$$

and so, in words, while the graph clustering coefficient  $C$  is the arithmetic average of the clustering coefficients of all the nodes, the transitivity  $T$  is the average of the clustering coefficients weighted by the number of triads centred on each node  $i$ .

In the following we will focus on the clustering coefficient  $C$ .

### B. Calculation of the clustering coefficient for some graphs

#### 1. Erdős-Rényi random graphs

Let us consider a graph picked up from the canonical ensemble  $G(p, N)$  composed by graphs with a fixed number of nodes  $N$  and a probability  $p$  to have an edge between two nodes. We want to

compute the expected clustering coefficient  $E[C]$  of such a graph (averaging over the ensemble) according to 3. We can consider a node  $i$  with degree  $k$ , its subgraph will contain in average  $k[G_i] = p \binom{k}{2}$  edges. Hence,  $c_i = p$ , and also the clustering coefficient is

$$C = p. \quad (5)$$

## 2. Random regular graphs

Let us consider a random  $d$ -regular graphs built according to the configuration model (cite bollobas paper). We want to calculate the expected clustering coefficient  $E[C]$  of such a graph (averaging over the ensemble). We are interested in

$$E[C] = \langle E[c] \rangle = \frac{1}{N} \sum_{i \in V} E[c_i], \quad (6)$$

so the main problem is calculating  $E[c_i] = \frac{E[K(G_i)]}{\binom{d}{2}}$

$$\begin{aligned} E[K(G_i)] &= \sum_{\langle jk \rangle} P(e_{jk} \in E, j \in \mathcal{N}(i), k \in \mathcal{N}(i), j \neq k) \\ &= \frac{1}{2} \sum_{j \in V \setminus i} P(j \in \mathcal{N}(i)) \sum_{k \in V \setminus i} P(k \in \mathcal{N}(i), k \neq j) P(e_{jk} \in E | j \in \mathcal{N}(i), k \in \mathcal{N}(i)) \\ &= \frac{1}{2} \sum_{j \in V \setminus i} P(j \in \mathcal{N}(i)) \sum_{k \in V \setminus i} P(k \in \mathcal{N}(i) | k \neq j) P(k \neq j) P(e_{jk} \in E | j \in \mathcal{N}(i), k \in \mathcal{N}(i)) \\ &= \frac{1}{2} \sum_{j \in V \setminus i} \frac{d}{N-1} \sum_{k \in V \setminus i} \frac{N-2}{N-1} \frac{d-1}{N-2} \frac{d-1}{N-1} \\ &= \frac{1}{2} \frac{d(d-1)^2}{N-1} \end{aligned} \quad (7)$$

where in the first line we consider summing over all the links  $\langle jk \rangle$  and in the following we rewrite the sum as considering to sum over the nodes instead of the links (easier). Finally we get

$$E[C] = \frac{E[K(G_i)]}{\binom{d}{2}} = \frac{d-1}{N-1} \quad (8)$$

## C. Numerical simulations to confirm the results

In the following we refer to the code contained in the file [Clustering th vs num.ipynb](#).

### 1. Erdős-Rényi random graphs

We consider Erdős-Rényi random graphs extracted from the usual canonical ensemble. We implemented the naive code in order to pick-up a graph from such an ensemble. The time complexity of the code is of  $O(N^2)$  as we can immediately see from the function `def ERG(p,N)`.

We run simulations for 6 different graph dimensions ( $N = 10, 50, 100, 500, 1000, 5000$ ) with average degree  $d = 4$  and for each dimension we average the clustering coefficient  $C$  over  $n = 10^2$  different graphs. We plot these values in fig. .... with their statistical errors (standard deviations from the mean value). On the same fig. .... we plot theoretical results we got from the expression in 5. We get the clustering coefficient out from the function `def C_numerical(edge list)`. The time complexity of such a function is of  $O(Nd(d-1)) = O(2K_{tot}(d-1))$  where  $K_{tot}$  is the total number of edges in the graph.

As we can see there is good agreement between theory and numerical simulations. Nevertheless the theory seems always to over-estimate the true value (why?).

### 2. Random regular graphs

Here we consider random  $d$ -regular graphs extracted from the configuration model (cite bollobas). The function we wrote to pick them up is `def RRG_direct(d,N)`. The algorithm follows this sequence of steps:

- (i) we begin with a set of  $N$  nodes;
- (ii) we create a set of  $Nd$  points and we distribute them across  $N$  buckets, such that each bucket contains  $d$  points;
- (iii) we take randomly a point and we pair it randomly with another one, until  $\frac{Nd}{2}$  pairs are obtained;
- (iv) we collapse the points, so that each bucket maps onto a single node of the former graph. We consider all edges between points as the edges of the corresponding nodes.
- (v) we check if the resulting graph is simple (no loops and multi-edges). If it is not we restart.

The time complexity of the algorithm to create the graph is of  $O(N^2)$  lead by (iv) and to check if it is good (v) it is  $O(Nd^2)$ .

Even here we run simulations for 6 different graph dimensions ( $N = 10, 50, 100, 500, 1000, 5000$ ) with degree  $d = 4$  and for each dimension we average the clustering coefficient  $C$  over  $n = 10^2$  different graphs. We plot these values in fig. .... with their statistical errors (standard deviations from the mean value). On the same fig. .... we plot theoretical results we got from the expression in 5.

As we can see there is good agreement between theory and numerical simulations. Nevertheless the theory seems always to over-estimate the true value (why?).

## II. RECIPROCATED ARCS IN DIRECTED REGULAR RANDOM GRAPHS

In a directed graph  $G(V, E)$  we say that the connection between node  $i$  and  $j$  is reciprocated if both  $(i, j) \in E$  and  $(j, i) \in E$ .

### A. Calculation of the expected fraction of reciprocated arcs in a directed random regular graph

We consider directed random  $d$ -regular graphs  $G(V, E)$  built according to the configuration model (a variant of the configuration model presented in the previous paragraph).

We are interested in finding an expression for the expected fraction of reciprocated arcs  $r$  in a graph picked up from such an ensemble.

The maximum number of reciprocated arcs is  $R_{max} = \frac{dN}{2}$  (i.e. the number of edges on an undirected random  $d$ -regular graph). We define  $r$  as follows:

$$r = \left\langle \frac{R}{R_{max}} \right\rangle \quad (9)$$

where  $R$  refers to the number of reciprocated arcs on a directed random  $d$ -regular graph and  $\langle \rangle$  refers to the average over the ensemble.

Finding  $r$  is then translated in calculating  $\langle R \rangle$ , we do it in the following way:

$$\begin{aligned} \langle R \rangle &= \frac{1}{2} \sum_{i \in V} \sum_{j \in V \setminus i} P(i \rightarrow j, j \rightarrow i) \\ &= \frac{1}{2} \sum_{i \in V} \sum_{j \in V \setminus i} P(i \rightarrow j) P(j \rightarrow i) \\ &= \frac{1}{2} \sum_{i \in V} \sum_{j \in V \setminus i} \frac{d}{N-1} \frac{d}{N-1} \\ &= \frac{1}{2} N(N-1) \frac{d^2}{(N-1)^2} \\ &= \frac{Nd^2}{2(N-1)}, \end{aligned} \quad (10)$$

where we exploited the fact that edges are independent random variables  $P(i \rightarrow j, j \rightarrow i) = P(i \rightarrow j)P(j \rightarrow i)$  and identically distributed too  $P(i \rightarrow j) = P(j \rightarrow i) = \frac{d}{N-1}$ .

We finally get that the expected fraction of reciprocated edges is

$$r = \left\langle \frac{R}{R_{max}} \right\rangle = \frac{d}{N-1} \quad (11)$$

### B. Numerical simulations to confirm the results