



CHORUS

This is the accepted manuscript made available via CHORUS. The article has been published as:

Cycles and clustering in multiplex networks

Gareth J. Baxter, Davide Cellai, Sergey N. Dorogovtsev, and José F. F. Mendes

Phys. Rev. E **94**, 062308 — Published 19 December 2016

DOI: [10.1103/PhysRevE.94.062308](https://doi.org/10.1103/PhysRevE.94.062308)

Cycles and Clustering in Multiplex Networks

Gareth J. Baxter,^{1,*} Davide Cellai,^{2,3} Sergey N. Dorogovtsev,^{1,4} and José F. F. Mendes¹

¹*Department of Physics & I3N, University of Aveiro, Portugal*

²*Idiro Analytics, Clarendon House, 39 Clarendon Street, Dublin 2, Ireland*

³*MACSI, Department of Mathematics and Statistics, University of Limerick, Ireland*

⁴*A. F. Ioffe Physico-Technical Institute, 194021 St. Petersburg, Russia*

(Dated: November 29, 2016)

In multiplex networks, cycles cannot be characterized only by their length, as edges may occur in different layers in different combinations. We define a classification of cycles by the number of edges in each layer and the number of switches between layers. We calculate the expected number of cycles of each type in the configuration model of a large sparse multiplex network. Our method accounts for the full degree distribution including correlations between degrees in different layers. In particular, we obtain the numbers of cycles of length 3 of all possible types. Using these, we give a complete set of clustering coefficients and their expected values. We show that correlations between the degrees of a vertex in different layers strongly affect the number of cycles of a given type, and the number of switches between layers. Both increase with assortative correlations and are strongly decreased by disassortative correlations. The effect of correlations on clustering coefficients is equally pronounced.

PACS numbers: 89.75.Hc, 89.75.Fb, 02.50.-r

I. INTRODUCTION

The realization that many complex systems cannot be understood by representing them as a single network, has led to an explosion of interest in multilayer and multiplex (multiple types of edges) networks. Applications range from infrastructure [1], financial [2], transport [3] and ecological [4].

To properly study multilayer systems, it is essential to understand the fundamental properties of such structures. Many concepts from single layer networks have already been generalized to multiple layers, from the degree distribution, to connectivity, adjacency and Laplacian matrices, centrality measures and so on [5–7]. In many cases, the generalization of concepts from single networks—for example, the meaning of “giant connected component” [8, 9]—is not straightforward, and introduces a new dimension to the problem.

In this Paper, we give an analytical description of the statistics of cycles in multiplex networks. In particular, we consider multiplex networks characterized by a given multi-degree distribution (configuration model). As this model is the starting point of any network analysis of a real-world system, it is easy to see how important it is to analytically characterize a structural property like the statistics of cycles. In a multilayer network, the possibility to switch between layers greatly increases the number of cycle types with respect to the single layer case. Cycles are then no longer defined simply by their length. We must take into account the proportion of the cycle in each layer, as well as the number of switches between layers. In particular, this leads to a set of different clus-

tering coefficients generated by different cycles of length 3.

Even when two cycles contain the same number of edges of each color, they can differ in the way colored edges are arranged within the cycle. We introduce a matrix \mathbf{s} characterizing the number of edges of each color, and the number of switches between layers of a cycle. We give formulæ for calculating the mean number of cycles corresponding to a given \mathbf{s} in a random graph with a given size and degree distribution. As examples, we calculate the distribution of edge colors and switches in cycles of a given length, show the effects of degree correlations between layers, and examine the special case of cycles of length 3, which allows the calculation of the generalized clustering coefficients in multiplex networks.

The statistics of cycles is relevant both from a theoretical and an applicative point of view. From a theoretical perspective, it allows one to understand whether the distribution of cycles observed in a real world network is significantly different from that in a random graph with similar statistics [10]. Even in the single layer case, the high concentration of finite cycles in real-world networks has been a formidable barrier to analytic treatment, as mathematical models of large networks are typically based on the local tree-like assumption, i.e. the vanishing of density of finite cycles as the size of the network diverges [11–13]. On the other hand, some analytic theories have been quite successful even in real-world networks [14], suggesting that the role played by the detailed architecture of topological correlations and cycles is not easily characterized. In multiplex networks, further types of correlations arise naturally, with implications for the structural properties of the multiplex [15, 16]. It is therefore necessary to turn now to multiplex cycles and investigate the statistics of each category based on edge color and layer switches, and the effects of correlations on them. The

* gjbaxter@ua.pt

properties of cycles is also an important topic of graph theory [17, 18]. A relatively smaller volume of work has been devoted to cycles on colored edge graphs, mainly involving only theorems of existence [19].

The statistics of cycles in multiplex networks is also relevant for a number of applications. In information technology, the presence of multiple paths of heterogeneous colors is a structural property that improves the robustness of a network [20] and the security of a wireless sensor network from a malicious attack [21]. Multiplex cycles are also relevant when examining commuter behavior on multiple transport networks, as they provide alternative routes [3]. In such applications, switching between layers may have a time or monetary cost, a consideration which is absent in single layer networks. The statistics of switches is therefore an essential part of any analysis of cycles in multiplex networks.

This paper is organized as follows. In Section II, we describe a classification of cycles in multiplex networks and explain our formula to calculate the number of cycles within a given class. In Section III, we use our formula to calculate multiplex clustering coefficients. Finally, in Section IV we state our conclusions.

II. STATISTICS OF CYCLES

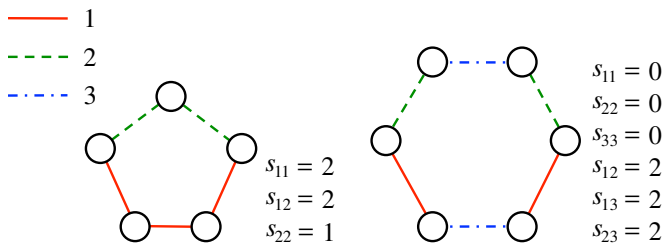


FIG. 1. Notation of multiplex cycles.

We characterize a given cycle in a multiplex network, by the matrix $\mathbf{s} = \{s_{ab}\}_{a,b=1,\dots,M}$, where each element s_{ab} defines the number of nodes in the cycle which connect an edge of type a with an edge of type b . When $a = b$, this counts the number of nodes where the cycle remains in the same layer. When $a \neq b$, s_{ab} counts the number of switches between layer a and layer b . In this Paper, we identify and study the classes of equivalence of cycles with the same \mathbf{s} . In addition, here we only consider cycles without an orientation, therefore the order of the switches is not important, i.e. $s_{ba} = s_{ab}$. Examples of cycles with a given \mathbf{s} are shown in Fig. 1. As a consequence of this definition, the total number of edges in layer a is then $n_a = s_{aa} + \frac{1}{2} \sum_{b \neq a} s_{ab}$. Clearly, the number of switches must be such that n_a is an integer for all a . The total length of a cycle L is the sum of all entries of \mathbf{s} .

We consider a generalization of the configuration model to multiplex networks, i.e. large sparse random

multiplex networks with N nodes in M layers, defined by the joint multidegree distribution $P(q_1, q_2, \dots, q_M)$. This ensemble includes all possible configurations with multidegree sequence sampled from this distribution with equal statistical weight [22, 23].

To calculate the mean number of cycles $\mathcal{N}(\mathbf{s})$ with a given \mathbf{s} in a random graph, we first count the number of ways we can select, from the given multidegree distribution, the nodes that have the connectivity required for each s_{ab} ; then, we count the number of ways we can connect these nodes to form a cycle. This method is similar to that used in, for example, Ref. [24] but we extend it to account for edges of different colors and switches between layers. The results can be written as the product of several factors:

$$\mathcal{N}(\mathbf{s}) = G(\mathbf{s})W(N, \mathbf{s})R[N, \mathbf{s}, P(\mathbf{q})], \quad (1)$$

where $R(N, \mathbf{s}, P(\mathbf{q}))$ counts the number of ways one can select pairs of edges connected to L nodes, in the correct numbers to match the elements of matrix \mathbf{s} , $W(N, \mathbf{s})$ counts the number of graphs in the ensemble containing the cycle, and $G(\mathbf{s})$ counts the number of ways of arranging the selected nodes to form a cycle. In principle one can complete this calculation for an arbitrary cycle in a multiplex with an arbitrary number of layers. However, the calculation of $G(\mathbf{s})$ becomes somewhat complicated for more than two layers, for anything but the shortest cycles.

Let us focus, now, on a two layer random multiplex (duplex), defined by the joint degree distribution $P(q_1, q_2)$. In the case of two layers, \mathbf{s} has three entries: s_{11} , s_{22} , and s_{12} . For a given L , s_{11} and s_{22} can take any value from 0 to L , while the number of switches $s_{12} \equiv 2p$, for integer p , to ensure that the number of edges of each color is integral, while all three must satisfy $s_{11} + s_{22} + s_{12} = L$. The formula (1) can be calculated explicitly in the asymptotic case characterized by $L \ll N$.

Without switches, $G(\mathbf{s})$ is simply the number of possible orderings of the L nodes, dividing by $2L$ as each direction and starting point in the ordering is equivalent:

$$G(L, 0, 0) = \frac{L!}{2L} \quad (2)$$

and similarly for $G(0, L, 0)$. When $s_{12} > 0$, $G(\mathbf{s})$ is given by the number of ways of ordering the switches $(s_{12} - 1)!$ multiplied by the number of ways $D(s_{11}, p)D(s_{22}, p)$ of placing the non-switching nodes in the spaces between the switches, where

$$\begin{aligned} D(s, p) &= \sum_{n_1=0}^l \binom{s}{n_1} n_1! \sum_{n_2=0}^{s-n_1} \binom{s-n_1}{n_2} n_2 \dots \\ &\dots \sum_{n_{p-1}=0}^{s-\sum_{j=1}^{p-2} n_j} \binom{s-\sum_{j=1}^{p-2} n_j}{n_{p-1}} n_{p-1}! \left[s - \sum_{j=1}^{p-1} n_j \right] \\ &= \frac{(s+p-1)!}{(p-1)!} \end{aligned} \quad (3)$$

Thus

$$G(s_{11}, s_{22}, s_{12}) = \begin{cases} (s_{12} - 1)! D(s_{11}, p) D(s_{22}, p) & s_{12} > 0 \\ \frac{(s_{11} + s_{22})!}{2^{(s_{11} + s_{22})}} & s_{12} = 0 \end{cases} \quad (4)$$

The number of ways to form layer 1 is the number of ways to connect $c_1 N$ stubs in pairs: $(c_1 N - 1)(c_1 N - 3) \dots 3 \cdot 1 = (c_1 N - 1)!!$, while the number of ways to connect the edges not forming part of the loop is $(c_1 N - 2s_{11} - s_{12} - 1)!!$. Hence the fraction of layer 1 configurations containing the loop is the ratio of these two numbers. Repeating for layer 2, we find that $W(N, \mathbf{s})$ can be simply written

$$W(N, \mathbf{s}) = \frac{(c_1 N - 2s_{11} - s_{12} - 1)!! (c_2 N - 2s_{22} - s_{12} - 1)!!}{(c_1 N - 1)!! (c_2 N - 1)!!} \quad (5)$$

A. Formulæ for short cycles

When $L \ll N$, the factor $W(N, \mathbf{s})$ can be approximated as

$$W(N, \mathbf{s}) = \frac{1}{N^L \langle q_1 \rangle^{s_{11} + p} \langle q_2 \rangle^{s_{22} + p}} \quad (6)$$

Furthermore, we can treat the selection of nodes as being done with replacement, meaning that $R(N, \mathbf{s}, P(\mathbf{q}))$ can be simply written as a product of terms for each node in the cycle:

$$R(N, \mathbf{s}, P(\mathbf{q})) = \frac{N^L \langle q_1 (q_1 - 1) \rangle^{s_{11}} \langle q_2 (q_2 - 1) \rangle^{s_{22}} \langle q_1 q_2 \rangle^{s_{12}}}{s_{11}! s_{22}! s_{12}!} \quad (7)$$

where $\langle \dots \rangle$ indicates averages with respect to the degree distribution (ensemble averages).

When Eq. (7) is combined with $G(\mathbf{s})$ and $W(N, \mathbf{s})$ as given by Eqs. (4) and (6) we find that, when none of s_{11}, s_{22}, s_{12} is zero,

$$\mathcal{N}(\mathbf{s}) = \binom{s_{11} + p - 1}{s_{11}} \binom{s_{22} + p - 1}{s_{22}} \frac{1}{2^p} \left[\frac{\langle q_1 (q_1 - 1) \rangle}{\langle q_1 \rangle} \right]^{s_{11}} \times \left[\frac{\langle q_2 (q_2 - 1) \rangle}{\langle q_2 \rangle} \right]^{s_{22}} \left[\frac{\langle q_1 q_2 \rangle}{\sqrt{\langle q_1 \rangle \langle q_2 \rangle}} \right]^{2p}, \quad (8)$$

In the case $s_{12} = 0$, the cycle consists of only one color, so $\mathcal{N}(\mathbf{s}) = 0$ unless either $s_{22} = 0$ or $s_{11} = 0$, in which case

$$\mathcal{N}(L, 0, 0) = \frac{1}{2^L} \left[\frac{\langle q_1 (q_1 - 1) \rangle}{\langle q_1 \rangle} \right]^L \quad (9)$$

This coincides with the single layer result found in, for example, Ref. [24]. Similarly, the formula for $\mathcal{N}(0, L, 0)$ is found simply by exchanging the subscripts.

On the other hand, when $s_{11} = 0$, it is still possible to have $s_{12} > 0$, when each segment in layer 1 consists

of only a single edge (i.e. each switch between layers is immediately followed by another switch). Then

$$\mathcal{N}(0, s_{22}, 2p) = \binom{s_{22} + p - 1}{s_{22}} \frac{1}{2^p} \left[\frac{\langle q_2 (q_2 - 1) \rangle}{\langle q_2 \rangle} \right]^{s_{22}} \left[\frac{\langle q_1 q_2 \rangle}{\sqrt{\langle q_1 \rangle \langle q_2 \rangle}} \right]^{2p}, \quad (10)$$

and similarly for the case $s_{22} = 0$ but $s_{12} > 0$ and $s_{11} > 0$, by exchanging the subscripts 1 and 2. If we project the two layers onto a single network, we recover the existing result for a single colored network, which has the same form as Eq. (9).

These results are valid for $L \ll N$, such as in the limit $N \rightarrow \infty$. The expected number of cycles in other cases, for longer cycles (when terms of $\mathcal{O}(1/N)$ can't be neglected) can be found by a more precise derivation, which we outline in the Appendix.

B. Representative examples

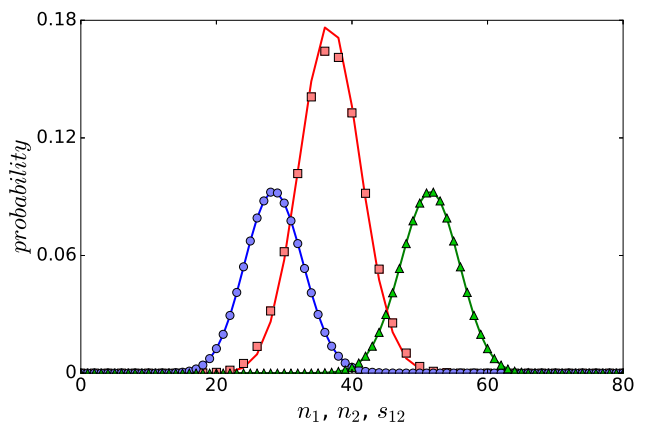


FIG. 2. Distribution of number of edges of type 1 (circles, blue online) and 2 (triangles, green online) and number of switches s_{12} (squares, red online) in cycles of length $L = 80$ in two uncorrelated layers with Poisson degree distributions having mean degrees $\langle q_1 \rangle = 25$ and $\langle q_2 \rangle = 45$. Symbols are from summation of Eq. (8), solid lines are binomial distributions for L trials with probabilities $\langle q_1 \rangle / (\langle q_1 \rangle + \langle q_2 \rangle)$, $\langle q_2 \rangle / (\langle q_1 \rangle + \langle q_2 \rangle)$, and $\langle q_1 \rangle \langle q_2 \rangle / (\langle q_1 \rangle + \langle q_2 \rangle)^2$ respectively.

The number of cycles having exactly $n_1 = s_{11} + \frac{1}{2}s_{12}$ edges of type 1 for a fixed L can be found by summing Eq. (8) over each s_{ab} . In the absence of inter-layer degree correlations, the resulting distributions for n_1 match the Binomial distribution found by selecting L edges at random from the network, as shown in Fig. 2. The mean number of type 1 edges is $p_1 = \langle q_1 \rangle / (\langle q_1 \rangle + \langle q_2 \rangle)$, and similarly for n_2 . In addition, the number of switches s_{12} , which must always be even, can be found by summing over s_{11} and s_{22} . The mean number of switches $\langle s_{12} \rangle$ is well predicted by $2L \langle q_1 \rangle \langle q_2 \rangle / (\langle q_1 \rangle + \langle q_2 \rangle)^2$, which is the expected

number of mismatches when pairing L randomly chosen edges, and the distribution is also well matched by a binomial distribution.

The number of cycles for a given matrix \mathbf{s} in Eq. (8) depends only on the moments of the joint degree distribution. This means that networks may have quite different degree distributions, and hence different structures, but if the relevant moments are the same, so will be the average number of cycles of each kind.

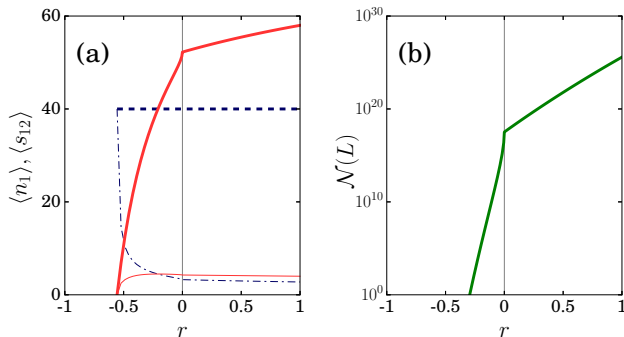


FIG. 3. (a) Mean for n_1 (blue, dashed) and s_{12} (red, heavy solid) and corresponding standard deviations (blue dash-dot, red light, respectively) as a function of Pearson correlation coefficient r for the degrees of a vertex in different layers. Assortative correlations ($r < 0$) are created using a joint degree distribution of the form $P(q_1, q_2) = \rho \frac{1}{2} [(P(q_1) + P(q_2))\delta_{q_1, q_2} + (1 - \rho)P(q_1)P(q_2)]$. disassortative correlations ($r < 0$) are of the form $P(q_1, q_2) = \rho [(P(q_1)\delta_{q_2, 0} + P(q_2)\delta_{q_1, 0}) + (1 - \rho)P(q_1)P(q_2)]$, where $P(q) \propto q^{-\gamma}$, with $\gamma = 3.7$, is a power-law distribution, and $L = 80$. Maximum disassortative correlation occurs when the two layers no longer overlap, at which point $n_1 = 0$. The corresponding value of r depends on γ . Figures are qualitatively similar for any value of $\gamma > 3$. (b) Total number of cycles as a function of r .

In simplex networks, it has been shown that degree-degree correlations for neighboring vertices affect quite deeply the number of cycles [25]. In multiplex networks, it is natural to ask about the effect of degree correlations across layers. Indeed, we can see from the last term in Eq. (8) that interlayer degree correlations affect the number of cycles having a given number of switches between layers. For a given L , assortative interlayer degree correlations will tend to increase the number of switches, as high degree nodes in one layer, which are more frequently visited, will also have more available edges in the other layers. Conversely, disassortative correlations will tend to decrease the number of switches. The mean number of edges of type 1 remains constant, although the distribution changes. These effects can be seen in Fig. 3, in which we plot the mean and standard deviation of n_1 and s_{12} as a function of the Pearson correlation coefficient r for the degrees of a vertex in different layers [11]. We see that in the extreme case of disassortative correlations, there are no switches and the entire cycle is of one color, as the standard deviation of n_1 reaches the maximum value of $L/2$. An even more dramatic change is

seen in the total number of cycles of a given length. The right panel in Fig. 3 shows the total number of cycles of length L as a function of r . Disassortative correlations greatly restrict the possible number cycles that can be formed, while assortative correlations greatly increase it. Note that the apparently sharp inflections at $r = 0$ result simply from the use of different functional forms for assortative and disassortative correlations.

III. CLUSTERING

The clustering coefficient of a network is related to the number of triangles, that is, cycles of length three. In such short cycles, the computation of factor $G(\mathbf{s})$ is straightforward, thus we can calculate the number of cycles of length 3 for any number of layers. Such a cycle may be entirely within one layer: $\mathcal{N}_m^{(1)} = z_m^3/6c_m^3$; have two edges in one layer (m) and one edge in a second layer (n): $\mathcal{N}_{m,n}^{(2)} = z_m \langle q_m q_n \rangle^2 / 2c_m^2 c_n$; or have one edge each in three different layers: $\mathcal{N}_{m,n,r}^{(3)} = \langle q_m q_n \rangle \langle q_m q_r \rangle \langle q_n q_r \rangle / c_m c_n c_r$, where $z_m \equiv \langle q_m (q_m - 1) \rangle$ and $c_m = \langle q_m \rangle$. We can then define a global clustering coefficient by

$$C = \frac{3 \sum_{\mathbf{s}: L=3} \mathcal{N}(\mathbf{s})}{\mathcal{V}(N)} = \frac{3 \sum_m \mathcal{N}_m^{(1)} + 3 \sum_{m,n \neq m} \mathcal{N}_{m,n}^{(2)} + 3 \sum_{\substack{m,n \neq m \\ r \neq m,n}} \mathcal{N}_{m,n,r}^{(3)}}{N \left[\sum_m z_m + 2 \sum_{m,n \neq m} \langle q_m q_n \rangle \right]}, \quad (11)$$

where $\mathcal{V}(N)$ is the number of adjacent edge-pairs in the graph, and the summation is over all cycle matrices \mathbf{s} having length $L = 3$.

One may also define partial clustering coefficients for triangles entirely within a given layer, two given layers, or three given layers ($C_{m,n,r}^{(3)}$) respectively:

$$C_m^{(1)} = \frac{3\mathcal{N}_m^{(1)}}{\frac{1}{2}Nz_m} = \frac{z_m^2}{Nc_m^3}, \quad (12)$$

$$C_{m,n}^{(2)} = \frac{3\mathcal{N}_{m,n}^{(2)}}{\frac{1}{2}N[z_m + 2\langle q_m q_n \rangle]} = \frac{3z_m \langle q_m q_n \rangle^2}{N[z_m + 2\langle q_m q_n \rangle] c_m^2 c_n}, \quad (13)$$

$$C_{m,n,r}^{(3)} = \frac{3\mathcal{N}_{m,n,r}^{(3)}}{\frac{1}{2}N[\langle q_m q_n \rangle + \langle q_m q_r \rangle + \langle q_n q_r \rangle]} \quad (14)$$

$$= \frac{6\langle q_m q_n \rangle \langle q_m q_r \rangle \langle q_n q_r \rangle}{Nc_m c_n c_r [\langle q_m q_n \rangle + \langle q_m q_r \rangle + \langle q_n q_r \rangle]}, \quad (15)$$

or triangles entirely in one, two, or three layers, regardless of which particular layers they are (these less fine-grained coefficients were defined and calculated numerically in

Ref. [10]):

$$C^{(1)} = \frac{\sum_m (z_m/c_m)^3}{N \sum_m z_m}, \quad (16)$$

$$C^{(2)} = \frac{3 \sum_{m,n \neq m} z_m \langle q_m q_n \rangle^2 / (c_m^2 c_n)}{N \sum_{m,n \neq m} [z_m + 2 \langle q_m q_n \rangle]}, \quad (17)$$

$$C^{(3)} = \frac{6 \sum_{m,n \neq m, r \neq m, n} \langle q_m q_n \rangle \langle q_m q_r \rangle \langle q_n q_r \rangle / c_m c_n c_r}{N \sum_{m,n \neq m, r \neq m, n} \langle q_m q_n \rangle}. \quad (18)$$

These formulæ give the expected clustering coefficients for large random graphs, taking into account the full degree distribution. This gives a more accurate result than found by simply matching the mean degree to an Erdős-Rényi network [10], for which the clustering coefficients can be calculated by considering the probability for a given edge to be present or absent:

$$C_{m(ER)}^{(1)} = \frac{c_m}{N}, \quad (19)$$

$$C_{m,n(ER)}^{(2)} = \frac{3c_m c_n}{N(c_m + 2c_n)}, \quad (20)$$

$$C_{m,n,r(ER)}^{(3)} = \frac{6c_m c_n c_r}{N(c_m c_n + c_m c_r + c_n c_r)}, \quad (21)$$

$$C_{ER}^{(1)} = \frac{\sum_m c_m^3}{N \sum_m c_m^2}, \quad (22)$$

$$C_{ER}^{(2)} = \frac{3 \sum_{m,n \neq m} c_m^2 c_n}{N \sum_{m,n \neq m} (c_m^2 + 2c_m c_n)}, \quad (23)$$

$$C_{ER}^{(3)} = \frac{6 \sum_{m,n \neq m, r \neq m, n} c_m c_n c_r}{N \sum_{m,n \neq m, r \neq m, n} c_m c_n}, \quad (24)$$

$$C_{ER} = \frac{\sum_m c_m}{2N}, \quad (25)$$

which coincide with the results found by inserting uncorrelated Poisson degree distributions into Eqs. (11)-(18).

To illustrate the importance of taking correlations into account, we compared our formulæ Eqs. (11)-(18) with measurements of synthetic networks. The results are summarized in Fig. 4. This shows that our method successfully accounts for the effects of broad degree distributions, and inter-layer correlations. In comparison, the Erdős-Rényi formulæ are generally only accurate for uncorrelated Erdős-Rényi layers and fail completely in the presence of strong correlations between layers.

IV. CONCLUSIONS

In single layer networks, cycles are characterized by their length. In multiplex networks, there are many more possibilities. In particular, there is the possibility to switch between layers, and this must be accounted for. In this Paper we have introduced a classification for cycles in multiplex networks based on the number of edges in each layer, and the number of switches between layers.

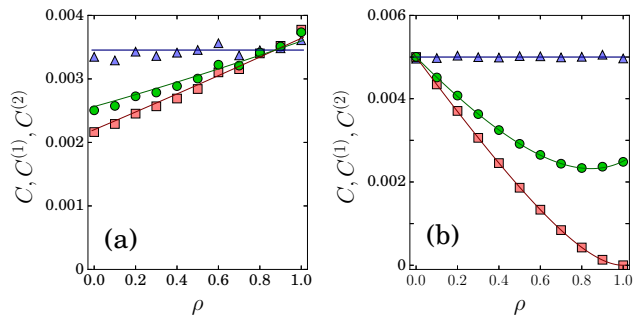


FIG. 4. (a) Total clustering coefficient, C (circles, green), single layer, $C^{(1)}$ (triangles, blue) and two layer, $C^{(2)}$ (squares, red), clustering coefficients for a two layer multiplex network with assortative inter layer degree correlations, joint degree distribution $P(q_1, q_2) = \rho^{1/2}[(P(q_1) + P(q_2))\delta_{q_1, q_2} + (1 - \rho)P(q_1)P(q_2)]$, where $P(q) \propto q^{-\gamma}$, with $\gamma = 2.9$ for $q \in [10, 100]$. Each layer has $N = 10^4$ vertices. For $\rho = 0$ correlations are absent, while for $\rho = 1$ degrees are perfectly correlated. Solid lines are expected theoretical values using Eqs. (11), (16), and (17). (b) Total C , single layer, $C^{(1)}$, and two layer, $C^{(2)}$, clustering coefficients for a two layer multiplex network with disassortative inter layer degree correlations, joint degree distribution $P(q_1, q_2) = \rho[(P(q_1)\delta_{q_2, 0} + P(q_2)\delta_{q_1, 0}) + (1 - \rho)P(q_1)P(q_2)]$, where $P(q)$ is a Poisson distribution with mean 50. For $\rho = 0$ correlations are absent, while for $\rho = 1$ degrees are perfectly disjoint. Solid lines are expected theoretical values using Eqs. (11), (16), and (17).

We further calculated the expected number of each type of cycle in a large random multiplex. Our results are valid for any multi-degree distribution, including distributions characterized by arbitrary degree correlations between layers. Interestingly, our formulæ show that the first and second order moments of the multi-degree distribution are sufficient to determine the statistics of cycles in large multiplex networks. The effect of correlations between a vertex's degrees in different layers affects these statistics through the degree-degree moment $\langle q_m q_n \rangle$. Assortative correlations tend to increase the number of switches in cycles of a given length, while also increasing the total number of cycles. Disassortative correlations, on the other hand, may greatly reduce both the total number of cycles and the total number of switches within these cycles.

These results further allow us to give the expected clustering coefficients in multiplex networks. The possibility that a closed triangle may have edges in multiple layers requires a more detailed discrimination of clustering coefficients. We give a complete classification of the various possible clustering coefficients and give their expected values. Inter layer correlations again have a strong effect, significantly increasing or decreasing the mixed-layer clustering coefficients. These results give a much more precise view of cycles and clustering in multiplex networks than using the mean degree alone, and establish the proper baseline for comparison with real-world

networks.

ACKNOWLEDGMENTS

We acknowledge useful discussions with Mikko Kivelä. This work was partially supported by the FET proactive IP project MULTIPLEX 317532. GJB was supported by the FCT grant No. SFRH/BPD/74040/2010. DC was supported by Science Foundation Ireland, Grant 14/IF/2461.

Appendix: The factor R for longer cycles

The derivation given in the main body of this Paper is valid for $L \ll N$. If one wants to consider longer cycles, a more complex derivation is required, which we sketch here.

1. Cycles in only one layer

For orientation, we first outline the derivation for cycles within a single layer, which is identical to that for a single network, and follows the method used in [24].

Consider selecting an arbitrary set of nodes from the network, of which n_q have degree q . There are $\binom{N_q}{n_q}$ ways to select n_q nodes, where $N_q = NP(q)$. We sum over all possible such sets and use a delta function to count only sets containing exactly L nodes. There are $q(q-1)$ ways to select the edges which connect a node of degree q in the cycle. Together these considerations mean we can write

$$R(N, L, P(q)) = N^L \sum_{\{n_q\}} \delta\left(\sum_q n_q - L\right) \prod_q \binom{N_q}{n_q} \left[\frac{q(q-1)}{N}\right]^{n_q}. \quad (\text{A.1})$$

We can write the delta function in integral form, and after replacing the sum over sets $\{n_q\}$ with a sum over n_q after the product, we have

$$\begin{aligned} R(N, L, P(q)) &= \frac{N^L}{2\pi} \int dx \prod_q \sum_{n_q=0}^{N_q} \binom{N_q}{n_q} \left[\frac{e^{ix} q(q-1)}{N}\right]^{n_q} e^{-ixL} \\ &= \frac{N^L}{2\pi} \int dx \prod_q \left[1 + \frac{e^{ix} q(q-1)}{N}\right]^{N_q} e^{-ixL}. \end{aligned} \quad (\text{A.2})$$

Writing the product as a sum within an exponential, and recognising that as $N_q = NP(q)$, this gives an average

over degree, we have

$$R(N, L, P(q)) = \frac{N^L}{2\pi} \int dx \exp\left\{-ixL + N \left\langle \log \left[1 + \frac{e^{ix} q(q-1)}{N}\right] \right\rangle\right\}. \quad (\text{A.3})$$

To evaluate this integral, we assume the integrand to be strongly peaked around a certain value x^* . Let

$$\phi(x) = \frac{-ixL}{N} + \left\langle \log \left[1 + \frac{e^{ix} q(q-1)}{N}\right] \right\rangle. \quad (\text{A.4})$$

A local maximum of this function occurs at x^* which is the solution of

$$\left\langle \frac{q(q-1)}{Ne^{-ix^*} + q(q-1)} \right\rangle = \frac{L}{N}. \quad (\text{A.5})$$

The second derivative evaluated at x^* is

$$\frac{d^2\phi}{dx^2}\Big|_{x=x^*} = - \left\langle \frac{q(q-1)}{[Ne^{-ix^*} + q(q-1)]^2} \right\rangle Ne^{-ix^*}. \quad (\text{A.6})$$

These results can be used to replace $\phi(x)$ with a Taylor expansion about x^* in Eq. (A.3). It has a simple Gaussian form that can easily be evaluated to give

$$R(N, L, P(q)) = \frac{N^L}{\sqrt{2\pi}} e^{N\phi(x^*)} \left\{ \left\langle \frac{q(q-1)}{[Ne^{-ix^*} + q(q-1)]^2} \right\rangle N^2 e^{-ix^*} \right\}^{-1/2}. \quad (\text{A.7})$$

To find an explicit expression, we must solve Eq. (A.5) and evaluate $\phi(x^*)$ and hence Eq. (A.7).

In the case $L \ll N$, Ne^{-ix^*} dominates $q(q-1)$, so that Eq. (A.5) becomes

$$\langle q(q-1) \rangle = Le^{-ix^*} \quad (\text{A.8})$$

so that

$$N\phi(x^*) = L + \log\left(\frac{\langle q(q-1) \rangle}{L}\right)^L \quad (\text{A.9})$$

and

$$\left\{ \left\langle \frac{q(q-1)}{[Ne^{-ix^*} + q(q-1)]^2} \right\rangle N^2 e^{-ix^*} \right\}^{-1/2} = \frac{1}{\sqrt{L}} \quad (\text{A.10})$$

giving

$$R(N, L, P(q)) = \frac{N^L}{\sqrt{2\pi L}} e^L \left[\frac{\langle q(q-1) \rangle}{L}\right]^L = \frac{N^L}{L!} \langle q(q-1) \rangle^L \quad (\text{A.11})$$

which when combined with Eqs. (2) and (6) gives exactly Eq. (9).

If instead we expand the lefthand side of Eq. (A.5) we find

$$\frac{s}{N} = \frac{\langle q(q-1) \rangle}{N e^{-ix^*}} - \frac{\langle q^2(q-1)^2 \rangle}{N^2 e^{-2ix^*}} + \dots \quad (\text{A.12})$$

and keeping the two leading terms, the solution is

$$e^{-ix^*} \approx \frac{\langle q(q-1) \rangle}{s} - \frac{\langle q^2(q-1)^2 \rangle}{N \langle q(q-1) \rangle} \quad (\text{A.13})$$

which can be substituted into Eqs. (A.4) and (A.7) to find the expected number of cycles when L is not small compared with N but $L \ll N^2$. The result will include dependence in N and on higher moments of the degree distribution.

2. Cycles in two layers

Now we outline the derivation for general cycles in a two layer multiplex. We proceed in the same way, but now there are three sets of nodes $\{n_{q_1, q_2}^{(1)}\}$ for the nodes connecting two edges in layer 1, $\{n_{q_1, q_2}^{(2)}\}$ for two edges in layer 2, and $\{n_{q_1, q_2}^{(s)}\}$ for the switches. We sum over all possible such sets, and use three delta functions to count only those whose total number of nodes (of all degrees) match s_{ij} . The combinatorial factor for the number of ways to select the three sets is

$$B_{n_1, n_2, n_3}^N = \frac{N!}{n_1! n_2! n_3! (N - n_1 - n_2 - n_3)!}. \quad (\text{A.14})$$

There are $q_1(q_1 - 1)$ ways to select the two edges connecting each of the $n_{q_1, q_2}^{(1)}$ nodes, and similarly for $n_{q_1, q_2}^{(2)}$, while for the $n_{q_1, q_2}^{(s)}$ switches the factor is $q_1 q_2$. These considerations give us

$$\begin{aligned} R(N, \mathbf{s}, P(\mathbf{q})) &= \sum_{\{n_{q_1, q_2}^{(1)}\}} \delta \left(\sum_{q_1, q_2} n_{q_1, q_2}^{(1)} - s_{11} \right) \\ &\sum_{\{n_{q_1, q_2}^{(2)}\}} \delta \left(\sum_{q_1, q_2} n_{q_1, q_2}^{(2)} - s_{22} \right) \sum_{\{n_{q_1, q_2}^{(s)}\}} \delta \left(\sum_{q_1, q_2} n_{q_1, q_2}^{(s)} - s_{12} \right) \\ &\prod_{q_1, q_2} B_{n_{q_1, q_2}^{(1)}, n_{q_1, q_2}^{(2)}, n_{q_1, q_2}^{(s)}}^{N_{q_1, q_2}} [q_1(q_1 - 1)]^{n_{q_1, q_2}^{(1)}} \\ &\times [q_2(q_2 - 1)]^{n_{q_1, q_2}^{(2)}} [q_1 q_2]^{n_{q_1, q_2}^{(s)}}. \quad (\text{A.15}) \end{aligned}$$

We can represent the delta function as an integral. Replacing the sums over sets $\{n_{q_1, q_2}^{(j)}\}$ with sums over $n_{q_1, q_2}^{(j)}$

after the product,

$$\begin{aligned} R(N, \mathbf{s}, P(\mathbf{q})) &= \frac{N^L}{(2\pi)^3} \int dx_1 dx_2 dx_3 \\ &\exp \left\{ ix_1 \left(\sum_{q_1, q_2} n_{q_1, q_2}^{(1)} - s_{11} \right) + ix_2 \left(\sum_{q_1, q_2} n_{q_1, q_2}^{(2)} - s_{22} \right) + \right. \\ &\quad \left. ix_3 \left(\sum_{q_1, q_2} n_{q_1, q_2}^{(s)} - s_{12} \right) \right\} \times \\ &\prod_{q_1, q_2} \sum_{n_{q_1, q_2}^{(1)}=0}^{N_{q_1, q_2}} \sum_{n_{q_1, q_2}^{(2)}=0}^{N_{q_1, q_2} - n_{q_1, q_2}^{(1)}} \sum_{n_{q_1, q_2}^{(s)}=0}^{N_{q_1, q_2} - n_{q_1, q_2}^{(1)} - n_{q_1, q_2}^{(2)}} B_{n_{q_1, q_2}^{(1)}, n_{q_1, q_2}^{(2)}, n_{q_1, q_2}^{(s)}}^{N_{q_1, q_2}} \\ &\left[\frac{q_1(q_1 - 1)}{N} \right]^{n_{q_1, q_2}^{(1)}} \left[\frac{q_2(q_2 - 1)}{N} \right]^{n_{q_1, q_2}^{(2)}} \left[\frac{q_1 q_2}{N} \right]^{n_{q_1, q_2}^{(s)}}. \quad (\text{A.16}) \end{aligned}$$

Each sum can be completed successively using the binomial identity. Finally converting the product into a sum within an exponential, and writing it as an expectation value we find

$$R(N, \mathbf{s}, P(\mathbf{q})) = \frac{N^L}{(2\pi)^3} \int dx_1 dx_2 dx_3 \exp \{ N \phi(\mathbf{x}) \}. \quad (\text{A.17})$$

where

$$\begin{aligned} \phi(\mathbf{x}) &= \frac{1}{N} \left[\log \left(\frac{y_1}{N} \right)^{s_{11}} + \log \left(\frac{y_2}{N} \right)^{s_{22}} + \log \left(\frac{y_3}{N} \right)^{s_{12}} \right] + \\ &\left\langle \log \left[1 + \frac{b_1}{y_1} + \frac{b_2}{y_2} + \frac{b_3}{y_3} \right] \right\rangle \quad (\text{A.18}) \end{aligned}$$

and $y_j = N e^{-ix_j}$ and $b_1 = q_1(q_1 - 1)$, $b_2 = q_2(q_2 - 1)$, and $b_3 = q_1 q_2$.

As before, we expand $\phi(\mathbf{x})$ about the local maximum \mathbf{x}^* which is the simultaneous solution of

$$\frac{s_{11}}{N} = \left\langle \frac{b_1}{y_1^* [1 + b_1/y_1^* + b_2/y_2^* + b_3/y_3^*]} \right\rangle \quad (\text{A.19})$$

$$\frac{s_{22}}{N} = \left\langle \frac{b_2}{y_2^* [1 + b_1/y_1^* + b_2/y_2^* + b_3/y_3^*]} \right\rangle \quad (\text{A.20})$$

$$\frac{s_{12}}{N} = \left\langle \frac{b_3}{y_3^* [1 + b_1/y_1^* + b_2/y_2^* + b_3/y_3^*]} \right\rangle. \quad (\text{A.21})$$

The Taylor expansion requires the second order derivatives:

$$\phi_{jj} \equiv \left. \frac{\partial^2 \phi}{\partial x_j^2} \right|_{\mathbf{x}^*} = -\frac{1}{y_j^*} \left\langle \frac{b_j (1 + b_k/y_k^* + b_l/y_l^*)}{(1 + b_1/y_1^* + b_2/y_2^* + b_3/y_3^*)^2} \right\rangle \quad (\text{A.22})$$

$$\phi_{jk} \equiv \left. \frac{\partial^2 \phi}{\partial x_j \partial x_k} \right|_{\mathbf{x}^*} = \frac{1}{y_j^* y_k^*} \left\langle \frac{b_j b_k}{(1 + b_1/y_1^* + b_2/y_2^* + b_3/y_3^*)^2} \right\rangle \quad (\text{A.23})$$

where the subscripts j, k, l each take one of the values 1, 2, 3. Substituting the second order Taylor expansion of $\phi(\mathbf{x})$ around \mathbf{x}^* into Eq. (A.17) yields a more complex

integral than in the single layer case, but its evaluation is still straightforward,

$$R(N, \mathbf{s}, P(\mathbf{q})) = \frac{N^L e^{N\phi(\mathbf{x}^*)}}{(2\pi N)^{3/2}} [\phi_{11}\phi_{22}\phi_{33} - \phi_{11}\phi_{23}^2 - \phi_{22}\phi_{13}^2 - \phi_{33}\phi_{12}^2 - 2\phi_{12}\phi_{13}\phi_{23}]^{-1/2}. \quad (\text{A.24})$$

To evaluate $R(N, \mathbf{s}, P(\mathbf{q}))$ for a given distribution, one

solves Eqs. (A.19)-(A.21), evaluates Eqs. (A.22) and (A.23) and substitutes into Eq. (A.24)).

For $L \ll N$ we can neglect $\mathcal{O}(1/N)$, giving $y_1^* = N\langle b_1 \rangle / s_{11}$ and similar expressions for y_2^* and y_3^* . The cross derivatives ϕ_{12} , ϕ_{13} , and ϕ_{23} vanish, and we recover Eq. (7). For larger L , dependence on N will remain, and there will be dependence on higher moments of the degree distribution.

-
- [1] S. M. Rinaldi, J. P. Peerenboom, and T. K. Kelly, "Identifying, understanding, and analyzing critical infrastructure interdependencies," *IEEE Control Syst. Mag.* **21**, 11–25 (2001).
- [2] X. Huang, I. Vodenska, S. Havlin, and H. E. Stanley, *Sci. Rep.* **3**, 1219 (2013).
- [3] M. De Domenico, A. Solé-Ribalta, S. Gómez, and A. Arenas, "Navigability of interconnected networks under random failures," *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8351–8356 (2014), arXiv:1306.0519.
- [4] M. J. O. Pocock, D. M. Evans, and J. Memmott, "The robustness and restoration of a network of ecological networks," *Science* **335**, 973–976 (2012).
- [5] S. Boccaletti, G. Bianconi, R. Criado, C. I. Del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, and M. Zanin, "The structure and dynamics of multilayer networks," *Phys. Rep.* **544**, 1–122 (2014).
- [6] M. Kivela, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, "Multilayer networks," *Journal of Complex Networks* **2**, 203–271 (2014).
- [7] M. De Domenico, C. Granell, M. A. Porter, and A. Arenas, "The physics of multilayer networks," arXiv preprint arXiv:1604.02021 (2016).
- [8] G. J. Baxter, S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, "Avalanche collapse of interdependent networks," *Phys. Rev. Lett.* **109**, 248701 (2012).
- [9] G. J. Baxter, D. Cellai, S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, "A unified approach to percolation processes on multiplex networks," in *Interconnected Networks* (Springer, 2016) pp. 101–123.
- [10] E. Cozzo, M. Kivela, M. De Domenico, A. Solé-Ribalta, A. Arenas, S. Gómez, M. A. Porter, and Y. Moreno, "Structure of triadic relations in multiplex networks," *New J. Phys.* **17**, 073029 (2015).
- [11] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review* **45**, 167–256 (2003), arXiv:cond-mat/0303516.
- [12] G. Bianconi and A. Capocci, "Number of loops of size h in growing scale-free networks," *Phys. Rev. Lett.* **90**, 078701 (2003).
- [13] B. Bollobás and O. M. Riordan, "Mathematical results on scale-free random graphs," in *Handbook of Graphs and Networks: From the Genome to the Internet*, edited by S. Bornholdt and H. G. Schuster (Wiley-VCH, Weinheim, 2003) pp. 1–34.
- [14] S. Melnik, A. Hackett, M. A. Porter, P. Mucha, and J. Gleeson, "The unreasonable effectiveness of tree-based theory for networks with clustering," *Phys. Rev. E* **83**, 036112 (2011).
- [15] J. Y. Kim and K. Goh, "Coevolution and correlated multiplexity in multiplex networks," *Phys. Rev. Lett.* **111**, 058702 (2013).
- [16] A. Hackett, D. Cellai, S. Gómez, A. Arenas, and J. P. Gleeson, "Bond percolation on multiplex networks," *Physical Review X* **6**, 021002 (2016).
- [17] Béla Bollobás, *Random Graphs*, 2nd ed. (Cambridge University Press, Cambridge, 2001).
- [18] F. Harary and B. Manvel, "On the number of cycles in a graph," *Math. Slovaca* **21**, 55–63 (1971).
- [19] A. Abouelaoualim, Das, W. Fernandez de la Vega, M. Karpinski, Y. Manoussakis, C. A. Martinhon, and R. Saad, "Cycles and paths in edge-colored graphs with given degrees," *J. Graph Theory* **64**, 63–86 (2010).
- [20] Y. Wang and Y. Desmedt, "Edge-colored graphs with applications to homogeneous faults," *Inform. Process. Lett.* **111**, 634–641 (2011).
- [21] K. Jain, "Security based on network topology against the wiretapping attack," *IEEE wirel. commun.* **11**, 68–71 (2004).
- [22] E. A. Bender and E. R. Canfield, "The asymptotic number of labeled graphs with given degree sequences," *J. Combinator. Theor. Series A* **24**, 296 (1978).
- [23] B. Bollobás, "A probabilistic proof of an asymptotic formula for the number of labelled regular graphs," *Eur. J. Combinator.* **1**, 311 (1980).
- [24] G. Bianconi and M. Marsili, "Loops of any size and hamilton cycles in random scale-free networks," *J. Stat. Mech. Theor. Exp.* **2005**, P06005 (2005).
- [25] G. Bianconi and M. Marsili, "Effect of degree correlations on the loop structure of scale-free networks," *Phys. Rev. E* **73**, 066127+ (2006).